Mauro Nisoli

Semiconductor Photonics

Principles and applications

SOCIETÀ EDITRICE
ESCULAPIO

Contents

Preface i

1 Band structure of semiconductors 1

1.1 Crystals, lattices and cells **1**

1.1.1 The Wigner-Seitz cell 6

1.2 The reciprocal lattice **7**

1.3 Electrons in a periodic crystal **10**

1.4 The concept of effective mass **13**

1.5 Energy bands **14**

1.5.1 Electrons and holes in a semiconductor 14

1.6 Calculation of the band structure **15**

1.6.1 Tight-binding method 15

1.6.2 Crystal with one-atom basis and single atomic orbital 18

1.6.3 Linear lattice 19

1.6.4 Simple cubic lattice 21

1.6.5 Band structure of semiconductors calculated by TBM 22

1.7 The $\mathbf{k} \cdot \mathbf{p}$ method **25**

1.8 Bandstructures of a few semiconductors **28**

1.8.1 Silicon 28

1.8.2 Gallium Arsenide 29

1.8.3 Gallium Nitride 30

2 Electrons in semiconductors 33

2.1 Introduction **33**

2.2 Density of states 33
2.2.1 Bulk (three-dimensional) semiconductors . . . 35
2.2.2 Two- and one-dimensional semiconductors . . . 36
2.3 Carrier statistics in semiconductors 37
2.3.1 Non-degenerate semiconductors . . . 40
2.4 Mass-Action Law 42
2.5 Doped semiconductors 43
2.5.1 Carrier density and Fermi energy in doped semiconductors . . . 47
2.5.2 Degenerate semiconductors . . . 48
2.6 Quasi-Fermi levels in non-equilibrium systems 49
2.7 Charge transport in semiconductors 52
2.8 Diffusion current 55

3 Basic concepts of quantum mechanics . . . 59
3.1 Quantum mechanics fundamentals 59
3.2 Time-dependent perturbation theory 64
3.2.1 Harmonic perturbation . . . 65
3.2.2 Transitions from a state to a band of states . . . 67
3.3 Properties of operators: a short summary 68
3.4 The density matrix 69
3.4.1 Temporal dependence of ρ . . . 73

4 Electron-photon interaction . . . 75
4.1 Introduction 75
4.2 Classical electromagnetic theory 75
4.3 Electrons in an electromagnetic field 78
4.4 Electric dipole approximation 81
4.5 Linear optical susceptibility 83
4.5.1 Relaxation processes . . . 84
4.5.2 Harmonic perturbation . . . 85
4.6 From optical susceptibility to absorption coefficient 88

5 Optical properties of semiconductors . . . 91
5.1 Stimulated transitions: selection rules 91
5.2 Joint density of states 92
5.3 Susceptibility and absorption coefficient in a semiconductor 94
5.3.1 Direct calculation of absorption coefficient . . . 98
5.4 Gain coefficient and Bernard-Duraffourg condition 99
5.5 Spontaneous emission 99
5.5.1 Bimolecular recombination . . . 104
5.6 Nonradiative recombination 107
5.6.1 Shockley-Read-Hall model . . . 107
5.6.2 Auger recombination . . . 111
5.6.3 Surface recombination . . . 117

5.7 Competition between radiative and nonradiative recombination 119

6 Quantum Wells 121
6.1 Introduction 121
6.2 Electronic states 122
6.2.1 Electronic states in the valence band 125
6.3 Density of states 127
6.4 Electron density 128
6.5 Transition selection rules 129
6.5.1 Intersubband transitions 131
6.6 Absorption and gain in a quantum well 132
6.7 Intersubband absorption 136
6.8 Strained quantum wells 139
6.9 Transparency density and differential gain 145
6.9.1 Differential gain 147
6.10 Excitons 148
6.10.1 Absorption spectrum 152
6.10.2 Excitons in quantum wells 154

7 Light Emitting Diodes 157
7.1 Basic concepts 157
7.2 Double heterostructure LEDs 160
7.3 Carrier leakage over barrier 165
7.4 External efficiency of a LED 168
7.5 Emission pattern of a LED 170
7.6 Luminous efficiency 171
7.7 Blue Light Emitting Diodes 173

8 Semiconductor Lasers 177
8.1 Introduction 177
8.2 Rate equations and threshold conditions for laser action 178
8.3 Temperature dependence 184
8.4 Output power 185
8.5 Quantum well lasers 186
8.6 Laser structures 189
8.7 Spectral and spatial characteristics of diode laser emission 191

9 Distributed Feedback Lasers 193
9.1 Basic concepts 193
9.2 Coupled-mode theory 196
9.2.1 Threshold conditions 199
9.3 DFB laser with uniform grating 201
9.3.1 Approximated evaluation of oscillating frequencies and threshold gain 203

9.4 DFB laser with $\lambda/4$-shifted grating 205
9.5 Distributed Bragg Reflector (DBR) laser 206

10 Vertical Cavity Surface-Emitting Lasers 209
10.1 Basic structure 209
10.2 Threshold conditions 210
10.3 Distributed Bragg reflectors for VCSELs 212
10.4 Threshold conditions and current confinement 215
10.5 Advantages and applications 217

11 Quantum cascade lasers 219
11.1 Quantum cascade lasers 219
11.2 Gain coefficient 221
11.3 Rate equations and threshold conditions 222
11.4 Output power, slope- and wall-plug efficiency 225
11.4.1 Wall-plug efficiency 226
11.5 Numerical example 229
11.6 Applications 230

A Fundamental constants 233

Index 235

Preface

The term *photonics* was introduced in the late sixties by Pierre Aigrain, who gave the following definition: "Photonics is the science of the harnessing of light. Photonics encompasses the generation of light, the detection of light, the management of light through guidance, manipulation and amplification and, most importantly, its utilisation for the benefit of mankind." In the last decades impressive progress in the field of photonics has been achieved, thanks to remarkable advances in the understanding of the physical processes at the heart of light-matter interaction in photonic applications and to the introduction of crucial technological innovations. Photonics is an extremely large field, as clearly pointed out by the above definition, since it refers to all types of technological device and process, where photons are involved.

This book does not aim to analyse all aspects of photonics: a few excellent textbooks already exist, which present several topics relevant for photonic applications. The aim of this book is to introduce and explain important physical processes at the heart of the optical properties of semiconductor devices, such as light emitting diodes (LEDs) and semiconductor lasers. It is suitable for a half-semester (or a one-semester) course in Photonics or Optoelectronics at the graduate level in engineering physics, electrical engineering or material science. It originated from a graduate course (Photonics I) which I am teaching at the Politecnico di Milano since 2006. The concepts of solid state physics and quantum mechanics which are required to understand the subjects discussed in this book are addressed in the introductory chapters. It is assumed that the reader has had courses on elementary quantum mechanics, solid-state physics, and electromagnetic theory at the undergraduate level.

The book presents a selection of topics, which I consider essential to understand the operation of semiconductor devices. It offers a relatively advanced analysis of the photo-physics of semiconductors, trying to avoid the use of exceedingly complex formalisms. Particular attention was devoted to offer a clear physical interpretation of all the obtained results. Various worked examples are added throughout all the

chapters to illustrate the application of the various formulas: the solved exercises are evidenced by the coloured boxes in the text. The numerical examples are also important since they allow the reader to have a direct feeling of the order of magnitude of the parameters used in the formulas discussed in the text. The grey boxes contain concise discussions of supplementary topics or more advanced derivations of particular results reported in the main text, which may not be easily derived by the reader.

Semiconductor Photonics is organized as follows. Chapter 1 focuses on the description of a few concepts of solid state physics, which are relevant for the calculation and analysis of the band structure of semiconductors. The Bloch theorem is introduced, which describes the wavefunction of electrons in periodic structures. The tight-binding method is considered, with a few simple examples, and the $\mathbf{k} \cdot \mathbf{p}$ method, which are used to calculated the band structure of semiconductors. Chapter 2 deals with the discussion of the main properties of charged particles (electrons and holes) in intrinsic and doped semiconductors. The density of states is first calculated and the essential concepts of carrier statistics in semiconductors are discussed. Basic concepts of quantum mechanics are contained in Chapter 3. In particular, the density matrix formalism is introduced, which is used in the book for the calculation of the optical susceptibility of a semiconductor. After a very short overview of essential aspects of classical electromagnetic theory, Chapter 4 analyses the interaction of electrons with an electromagnetic field. The expressions of the interaction Hamiltonian, which are extensively used throughout the book, are derived in this chapter. Chapters 5 and 6 build on the previous chapters. In particular, Chapter 5 deals with the optical properties of bulk semiconductors, i.e., semiconductors with spatial dimensions much larger than the de Broglie wavelength of the electrons involved in the relevant physical processes. Absorption and gain coefficients are calculated and the radiative and non-radiative recombination processes in semiconductors are analysed. Chapter 6 analyses the principles of the photo-physics in semiconductor quantum wells, i.e., in semiconductor structures where the electrons are confined in one direction by a potential well, with a thickness smaller than the electron de Broglie wavelength.

In the remaining five chapters the general results obtained in the first part of the book are applied to the investigation of the main optical properties of semiconductor devices: light-emitting diodes and lasers. The general philosophy adopted in these chapters is the following: the fundamental physical processes are investigated, rather than the technological characteristics of the devices. After a short and general analysis of semiconductor lasers in Chapter 8, based on the rate equation approach, Chapter 9 contains a detailed theoretical analysis of the Distributed Feedback (DFB) lasers, based on the use of the coupled-mode equations. By using a simple perturbative approach, the threshold laser conditions are obtained. Vertical Cavity Surface Emitting Lasers (VCSELs) and Quantum Cascade Lasers are analysed in the final two chapters.

Final note: why Escher's "*Sky and Water I*" on the book cover? This print suggested me various connections with photonics in semiconductors. It is light which, by playing with shapes, gives rise to an evolution which transforms fishes in water into birds in the sky and the transformation is closely related to the reciprocal interaction among elements. In a similar manner, it is light which, by playing with electrons, triggers dynamical processes inside matter, with the generation of new properties and functions, where complex interactions among particles and surrounding environment play a crucial role. The analogy can be pushed even more considering

that in Escher's print the transition from sky to water starts and ends with realistic shapes of birds and fishes, respectively, and moves from one real shape to the other through a sequence of shapes that, in particular in the central portion of the print, resemble but are not exactly neither birds nor fishes. Forcing a little the analogy, we can say that typical photonic processes in semiconductors evolve from a given physical (real) state to another physical (real) state through a sequence of quantum superposition of states, which do not have an analogy in the realm of classical physics.

Milano, 11 November 2016 *Mauro Nisoli*

1. Band structure of semiconductors

1.1 Crystals, lattices and cells

A crystal is composed by a periodic repetition of identical groups of atoms: a group is called *basis*. The corresponding crystal lattice is obtained by replacing each group of atoms by a representative point, as shown in Fig. 1.1. A crystal can be also called *lattice with a basis*. When the basis is composed by a single atom the corresponding lattice is called monoatomic. In a *Bravais lattice* the position $\mathbf{R}$ of all points in the lattice can be written as :

$$\mathbf{R} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3, \tag{1.1}$$

where $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$ are three non-coplanar translation vectors, called *primitive vectors* and n_1, n_2 and n_3 are arbitrary (positive or negative) integers. This definition of Bravais lattice implies that this lattice looks exactly the same when viewed from any lattice point. Not only the arrangement of points but also the orientation must be exactly the same from every point in a Bravais lattice. Therefore, two points in

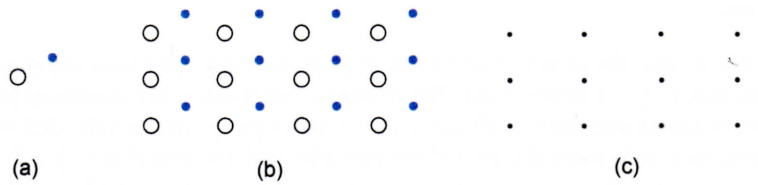

Figure 1.1: (a) Basis composed by two different atoms; (b) bidimensional crystal and (c) corresponding lattice.

the lattice, whose positions vectors are given by $\mathbf{r}$ and $\mathbf{r}' = \mathbf{r} + \mathbf{R}$ are completely equivalent environmentally. For example, the two-dimensional honeycomb lattice

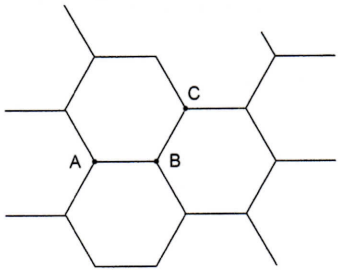

Figure 1.2: A two-dimensional honeycomb lattice is not a Bravais lattice.

shown in Fig. 1.2 is not a Bravais lattice. Indeed the lattice looks the same when it is viewed from points A and C, but not when it is viewed from point B: in this case the lattice appears rotated by 180°. We note that for a given Bravais lattice, the choice of the primitive vectors is not unique.

A lattice can be constructed by infinite repetitions, by translations, of a single cell without any overlapping. This cell can be primitive or non-primitive (or conventional). Primitive and conventional cells are not uniquely determined, as clearly illustrated in Fig. 1.3. The primitive cell has the minimum possible volume, given by:

$$V_u = \mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3), \tag{1.2}$$

and contains exactly one lattice point. Therefore in Fig. 1.3, which refers to a two-

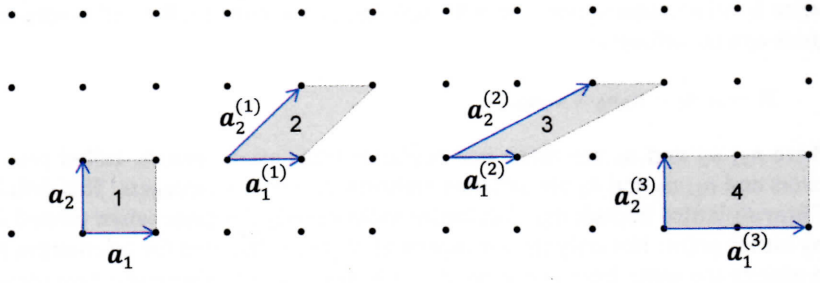

Figure 1.3: Two-dimensional lattice: cells 1, 2 and 3 are primitive cells. Cell 4 is not primitive.

dimensional case, the cells 1, 2 and 3 are all primitive cells: they have the same area and contains $4 \cdot \frac{1}{4} = 1$ lattice point. The primitive cell contains the minimum possible number of atoms and there is always a lattice point per primitive cell. Cell 4 is not primitive: its area is twice the area of the primitive cell and contains $4 \cdot \frac{1}{4} + 2 \cdot \frac{1}{2} = 2$ lattice points. Not all points in the lattice are linear combinations of $\mathbf{a}_1^{(3)}$ and $\mathbf{a}_2^{(3)}$ with integral coefficients. In some cases it is more convenient to consider conventional unit cells, with larger volumes (integer multiple of that of the primitive cell) but characterized by the same symmetry of the lattice.

Without entering into any detail about group theory, we can say that all the possible lattice structures are determined by the symmetry group which describes their

properties. A lattice structure can be transformed into itself not only by the translations 1.1, which define the translational group, but also by many other symmetry operations. The symmetry operations transforming a lattice into itself keeping at least one point fixed form a group called the *point group*. In the case of three-dimensional structures, the point symmetry gives rise to 14 types of lattices, which can be classified depending on the relationships between the amplitudes of the vectors $\mathbf{a}_i$ and the angles α, β and γ between them. As reported in Table 1.1, the 14 types of lattices can be grouped in one triclinic, two monoclinic, four orthorhombic, two tetragonal, three cubic, one trigonal and one hexagonal lattices.

Many semiconductors are characterized by a cubic lattice or by an hexagonal

System	Number of lattices	Amplitudes a_i and angles
Triclinic	1	$a_1 \neq a_2 \neq a_3$ $\alpha \neq \beta \neq \gamma$
Monoclinic	2	$a_1 \neq a_2 \neq a_3$ $\alpha = \gamma = 90° \; \beta \neq 90°$
Orthorhombic	4	$a_1 \neq a_2 \neq a_3$ $\alpha = \beta = \gamma = 90°$
Tetragonal	2	$a_1 = a_2 \neq a_3$ $\alpha = \beta = \gamma = 90°$
Cubic	3	$a_1 = a_2 = a_3$ $\alpha = \beta = \gamma = 90°$
Trigonal	1	$a_1 = a_2 = a_3$ $\alpha = \beta = \gamma < 120°, \neq 90°$
Hexagonal	1	$a_1 = a_2 \neq a_3$ $\alpha = \beta = 90°, \; \gamma = 120°$

Table 1.1: The 14 lattice systems in three dimensions (the last column shows the amplitudes a_i and the angles between vectors $\mathbf{a}_i$ of the unit cell).

lattice. There are three types of cubic lattices: simple cubic (sc), body-centered cubic (bcc) and face-centered cubic (fcc), whose unit cells are shown in Fig. 1.4. Note that

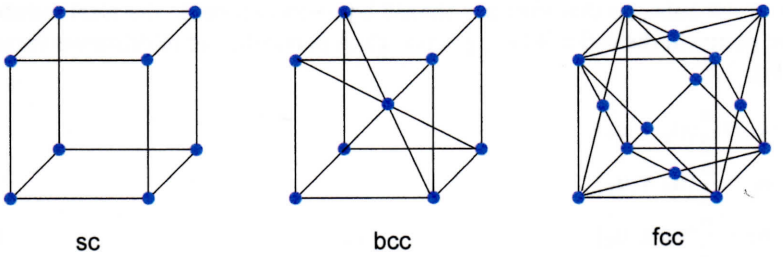

Figure 1.4: Cubic lattices: (sc) simple cubic; (bcc) body-centered cubic; (fcc) face-centered cubic.

only the simple cubic is a primitive cell, with volume a^3 and one lattice point per cell ($8 \times \frac{1}{8}$). The body-centered cubic lattice can be obtained from the simple cubic by placing a lattice point at the center of the cube. The conventional cell is the cube with edge a. It has 2 lattice points per unit cell ($8 \times \frac{1}{8} + 1$). In terms of cube edge a,

a set of primitive vectors can be written as:

$$\begin{aligned}\mathbf{a}_1 &= \frac{a}{2}(\mathbf{u}_x + \mathbf{u}_y - \mathbf{u}_z)\\ \mathbf{a}_2 &= \frac{a}{2}(-\mathbf{u}_x + \mathbf{u}_y + \mathbf{u}_z)\\ \mathbf{a}_3 &= \frac{a}{2}(\mathbf{u}_x - \mathbf{u}_y + \mathbf{u}_z)\end{aligned} \tag{1.3}$$

where $\mathbf{u}_x$, $\mathbf{u}_y$ and $\mathbf{u}_z$ are the unit vectors of the x, y and z axis, as shown in Fig. 1.5. The corresponding primitive cell, with volume $a^3/2$, contains by definition only one lattice point. This primitive cell does not have an obvious relation with the point

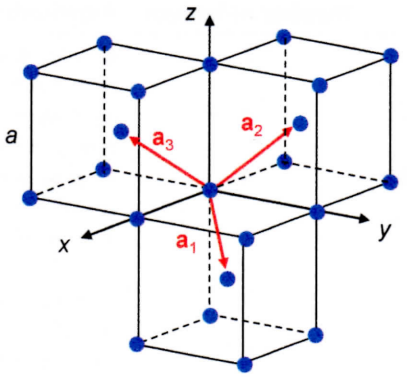

Figure 1.5: Set of primitive vectors for a body-centered cubic lattice.

symmetry (cubic) of the lattice. For this reason it is useful to consider a unit cell larger than the primitive cell and with the same symmetry of the crystal. For a bcc lattice the unit cell is a cube with edge a, with a volume which is twice the volume of the primitive cell. The bcc lattice can be also considered as a sc lattice with a two-point basis $0, (a/2)(\mathbf{u}_x + \mathbf{u}_y + \mathbf{u}_z)$.

The face-centered cubic Bravais lattice can be obtained from the sc lattice by adding a point in the center of each face. The fcc structure has lattice points on the faces of the cube, so that they are shared between two cells: the total number of lattice points in the cell is 4 ($8 \times \frac{1}{8} + 6 \times \frac{1}{2}$). A particular set of primitive vectors is (see Fig. 1.6):

$$\begin{aligned}\mathbf{a}_1 &= \frac{a}{2}(\mathbf{u}_y + \mathbf{u}_z)\\ \mathbf{a}_2 &= \frac{a}{2}(\mathbf{u}_z + \mathbf{u}_x)\\ \mathbf{a}_3 &= \frac{a}{2}(\mathbf{u}_x + \mathbf{u}_y)\end{aligned} \tag{1.4}$$

The primitive cell has a volume of $a^3/4$ and contains one lattice point. Also in this case the unit cell is generally assumed as a cube of edge a, with a volume which is four times the volume of a primitive cell. The fcc can be described as a sc lattice with a four-point basis $0, (a/2)(\mathbf{u}_x + \mathbf{u}_y), (a/2)(\mathbf{u}_y + \mathbf{u}_z), (a/2)(\mathbf{u}_z + \mathbf{u}_x)$. We recall that the numbers giving the size of the unit cell (for example, the number a in the case of a cubic crystal) are called lattice constants.

Many important semiconductors, for example silicon and germanium have a

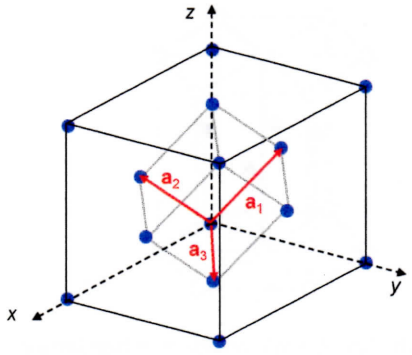

Figure 1.6: Primitive vectors and primitive cell of the face-centered cubic lattice.

diamond structure, which is the lattice formed by the carbon atoms in a diamond crystal. This structure consists of identical atoms, which occupy the lattice points of two inter-penetrating fcc lattices, which are displaced from each other along the body diagonal of the cubic cell by one quarter the length of the diagonal, as shown in Fig. 1.7. It can be seen as a fcc lattice with a two-point basis $0, (a/4)(\mathbf{u}_x + \mathbf{u}_y + \mathbf{u}_z)$. The

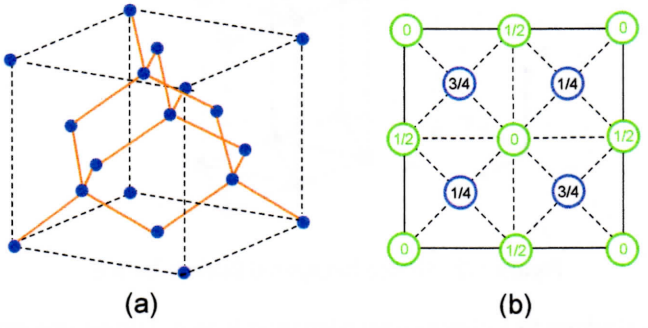

Figure 1.7: (a) Crystal structure of diamond; the solid lines shows the tetrahedral bond geometry; (b) atomic position in the cubic cell projected on a cube face. The fractions correspond to the height above the cube face in unit of cube edge a; the two colors correspond to the two inter-penetrating fcc lattices, which generate the diamond structure.

four nearest neighbours of each point are on the vertices of a regular tetrahedron. Note that the diamond lattice is not a Bravais lattice, since it does not look exactly the same when it is viewed from two nearest neighbour points. Since the unit cell of a fcc structure contains 4 lattice points, the unit cell of the diamond structure contains 8 lattice points. In this case it is not possible to choose a primitive cell in such a way that the basis of diamond contains only one atom. When the atoms which occupy one of the two fcc structures are different from the atoms occupying the other, the structure is called *zincblende* structure. Several semiconductors are characterized by this structure, such as GaAs, AlAs and many others.

Largely used semiconductors such as GaN, AlN, BN, SiC, have a hexagonal close-packet (hcp) structure (wurtzite structure). Also the hcp lattice is not a Bravais lattice. This lattice can be seen as two inter-penetrating simple hexagonal Bravais lattices,

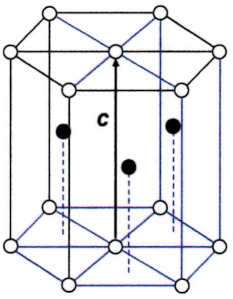

Figure 1.8: Atomic position in hcp lattice.

displaced vertically by $c/2$ in the direction of the common $c-$axis, and displaced in the horizontal plane in such a way that the points of one simple hexagonal lattice are placed above the centres of the triangles formed by the points of the other simple hexagonal lattice. Figure 1.9 shows a simple hexagonal Bravais lattice, which is

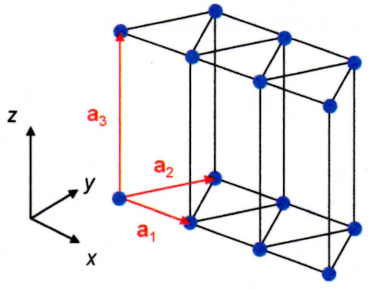

Figure 1.9: Simple hexagonal Bravais lattice.

obtained by stacking two-dimensional triangular Bravais lattices one exactly above the other along a direction perpendicular to each two-dimensional lattice. This stacking direction is usually called crystallographic $c-$axis. The primitive vectors can be written as:

$$\begin{aligned} \mathbf{a}_1 &= a\mathbf{u}_x \\ \mathbf{a}_2 &= \frac{a}{2}\mathbf{u}_x + \frac{\sqrt{3}}{2}a\mathbf{u}_y \\ \mathbf{a}_3 &= c\mathbf{u}_z. \end{aligned} \tag{1.5}$$

1.1.1 The Wigner-Seitz cell

The primitive cell can be also chosen in such a way that it presents the full symmetry of the Bravais lattice. This can be achieved by considering the Wigner-Seitz cell. The mathematical definition is the following: the Wigner-Seitz cell around a given lattice point is the spatial region that is closest to that particular lattice point than to any of other lattice points. It can also be demonstrated that the Wigner-Seitz cell is a primitive cell. The above definition does not refer to any particular choice of primitive vectors, for this reason the Wigner-Seitz cell is as symmetrical as the

Bravais lattice. The procedure for the construction of a Wigner-Seitz cell can be illustrated in the simple case of a two-dimensional rectangular lattice, as shown in Fig. 1.10(a). To determine the Wigner-Seitz cell about the lattice point P we have

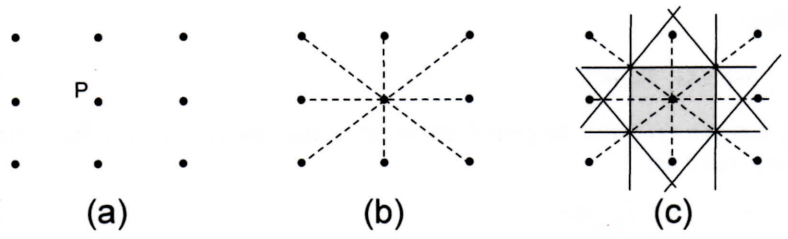

Figure 1.10: Construction of the Wigner-Seitz cell of a two-dimensional rectangular lattice.

first to draw the lines from P to all of its nearest neighbours (Fig. 1.10(b)) and then the bisectors to each of these lines (Fig. 1.10(c)). The Wigner-Seitz cell is the innermost region bounded by the perpendicular bisectors, as shown by the shaded region in Fig. 1.10(c). The same procedure can be applied in the case of a generic three-dimensional lattice.

1.2 The reciprocal lattice

In order to develop an analytic study of a crystalline solid, it is often useful to introduce the concept of reciprocal lattice, which basically represents the Fourier transform of the Bravais lattice. Assuming a Bravais lattice defined by the primitive translation vectors $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$, the reciprocal lattice can be defined by introducing its primitive translation vectors $(\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3)$ in analogy with the lattice in real space. The axis vectors of the reciprocal space can be written as:

$$\begin{aligned}
\mathbf{b}_1 &= \frac{2\pi}{V_u}\, \mathbf{a}_2 \times \mathbf{a}_3 \\
\mathbf{b}_2 &= \frac{2\pi}{V_u}\, \mathbf{a}_3 \times \mathbf{a}_1 \\
\mathbf{b}_3 &= \frac{2\pi}{V_u}\, \mathbf{a}_1 \times \mathbf{a}_2
\end{aligned} \tag{1.6}$$

where V_u is the volume of the unit cell given by:

$$V_u = \mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3). \tag{1.7}$$

The reciprocal lattice can be mapped by using the general translation vector $\mathbf{G}$ given by:

$$\mathbf{G} = m_1\mathbf{b}_1 + m_2\mathbf{b}_2 + m_3\mathbf{b}_3, \tag{1.8}$$

where m_1, m_2 and m_3 are integers. Vector $\mathbf{G}$ is called reciprocal lattice vector. Note that each vector given by 1.6 is orthogonal to two axis vectors of the crystal lattice, so that:

$$\mathbf{b}_i \cdot \mathbf{a}_j = 2\pi\delta_{ij}, \tag{1.9}$$

where δ_{ij} is the Kronecker delta symbol: $\delta_{ij} = 0$ for $i \neq j$ and $\delta_{ij} = 1$ for $i = j$. Moreover:

$$\mathbf{G} \cdot \mathbf{R} = 2\pi(n_1 m_1 + n_2 m_2 + n_3 m_3) = 2\pi\ell, \qquad \ell = 0, \pm 1, \pm 2, \ldots \tag{1.10}$$

so that:

$$\exp(i\ \mathbf{G} \cdot \mathbf{R}) = 1 \tag{1.11}$$

Any function $f(\mathbf{r})$ with the periodicity of the crystal lattice, i.e., $f(\mathbf{r} + \mathbf{R}) = f(\mathbf{r})$, can be expanded as:

$$f(\mathbf{r}) = \sum_{\mathbf{G}} f_{\mathbf{G}}\ e^{i\mathbf{G}\cdot\mathbf{r}} \tag{1.12}$$

where:

$$f_{\mathbf{G}} = \frac{1}{V_u} \int_{cell} f(\mathbf{r})\ e^{-i\mathbf{G}\cdot\mathbf{r}}\ d\mathbf{r}. \tag{1.13}$$

Indeed:

$$f(\mathbf{r} + \mathbf{R}) = \sum_{\mathbf{G}} f_{\mathbf{G}}\ e^{i\mathbf{G}\cdot(\mathbf{r}+\mathbf{R})} = f(\mathbf{r}), \tag{1.14}$$

where we have used Eq. 1.11.
While vectors in the direct lattice have the dimension of length, the vectors in the reciprocal lattice have the dimension of [length]$^{-1}$. The reciprocal space is therefore the most convenient space for wave vectors $\mathbf{k}$. Since each point in the reciprocal space can be reached by the translation vector $\mathbf{G}$, it is evident that we can restrict our analysis to a unit cell defined by the vectors $\mathbf{b}_i$. The Wigner-Seitz cell of the reciprocal lattice is called the *first Brillouin zone*. There are also the second, third, etc. Brillouin zones, at increasing distance from the origin, all with the same volume. These higher-order Brillouin zones can be translated into the first zone by adding suitable translation vectors $\mathbf{G}$. A Wigner-Seitz cell in the reciprocal space can be constructed following the same procedure used in the real space. For example, in the simple case of a linear lattice, we have a single primitive vector in the direct space: $\mathbf{a}_1 = a\mathbf{u}_x$. From Eq. 1.9 it is obvious that the primitive vector in the reciprocal space is:

$$\mathbf{b}_1 = \frac{2\pi}{a}\mathbf{u}_x. \tag{1.15}$$

The corresponding translation vector is given by:

$$\mathbf{G} = m\frac{2\pi}{a}\mathbf{u}_x, \qquad m = \pm 1, \pm 2, \ldots \tag{1.16}$$

Following the rules for the construction of the first Brillouin zone, we have that this region extends from $k = -\pi/a$ to $k = \pi/a$.
Following the same procedure it is possible to construct the first Brillouin zone for three-dimensional structures. For a simple cubic lattice, where:

$$\mathbf{a}_1 = a\mathbf{u}_x,\ \mathbf{a}_2 = a\mathbf{u}_y,\ \mathbf{a}_3 = a\mathbf{u}_z, \tag{1.17}$$

by using Eq. 1.6, where $V_u = a^3$:

$$\mathbf{b}_1 = \frac{2\pi}{a}\mathbf{u}_x,\ \mathbf{b}_2 = \frac{2\pi}{a}\mathbf{u}_y,\ \mathbf{b}_3 = \frac{2\pi}{a}\mathbf{u}_z, \tag{1.18}$$

Therefore the reciprocal lattice of a simple-cubic Bravais lattice with cubic primitive cell of edge a is a simple cubic lattice with cubic primitive cell of edge $2\pi/a$ and the first Brillouin zone is defined as follows:

$$-\frac{\pi}{a} \leq k_i \leq \frac{\pi}{a}, \qquad i = x, y, z. \tag{1.19}$$

This is shown in Fig. 1.11(a). The center of the first Brillouin zone is always called

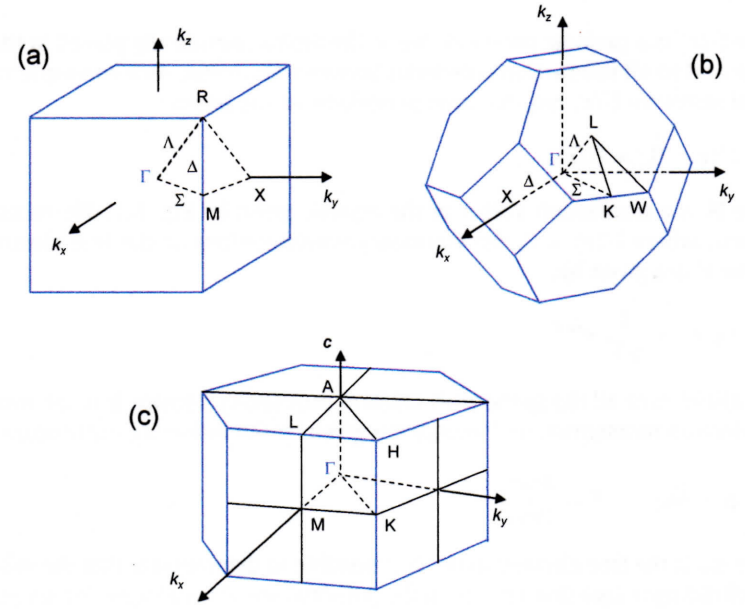

Figure 1.11: First Brillouin zone for (a) the simple cubic lattice, (b) the fcc lattice which forms the underlying Bravais lattice for diamond and zinc–blende structures and (c) the hexagonal wurtzite structure. Reported are the names of points and directions of high symmetry.

the $\Gamma-$point. A typical convention is to call high-symmetry points and directions inside the Brillouin zone by greek letters and high-symmetry points on the surfaces of the Brillouin zone by roman letters. For example, in the case of the fcc structure (see Fig. 1.11(b)) the three high-symmetry directions [100], [110] and [111] are denoted by:

[100] direction : $\dot{\Gamma}\ \overline{\Delta}\ \dot{X}$

[111] direction : $\dot{\Gamma}\ \overline{\Lambda}\ \dot{L}$

[110] direction : $\dot{\Gamma}\ \overline{\Sigma}\ \dot{K}$

The $X-$point at $(2\pi/a)(1,0,0)$ identifies the zone edge along the six equivalent [100] directions. The $L-$point is at $(\pi/a)(1,1,1)$ and it is at the zone edge along the eight equivalent [111] directions.

1.3 Electrons in a periodic crystal

After this brief overview of introductory concepts of solid state physics, we will study the interactions of electrons with a periodic structure. We will consider bulk semiconductors, i.e., semiconductors with spatial dimensions much larger than the de Broglie wavelength ($\lambda_B = h/p$, where p is the particle momentum) of the electrons involved in the interaction. In order to study the electronic and optical properties of the crystal, we have first to calculate the electronic wavefunctions and their energies inside the crystal. The system is described by the following Schrödinger equation:

$$\left[-\frac{\hbar^2}{2m_0}\nabla^2 + U(\mathbf{r})\right]\psi(\mathbf{r}) = \mathcal{E}\psi(\mathbf{r}) \tag{1.20}$$

where $U(\mathbf{r})$ is a periodic potential due to the atoms periodically placed in the crystal lattice and to all interaction potentials between electrons. As a consequence of the crystal structure $U(\mathbf{r})$ has the same periodicity of the lattice:

$$U(\mathbf{r}+\mathbf{R}) = U(\mathbf{r}), \tag{1.21}$$

where $\mathbf{R}$ is a translation vector of the crystal, given by Eq. 1.1. We recall that in vacuum, where $U(\mathbf{r}) = 0$, the stationary wavefunctions of the free electrons in a volume V are given by:

$$\psi_{\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{V}}\, e^{i\mathbf{k}\cdot\mathbf{r}}$$

delocalized over all the space with uniform probability density. $\mathbf{k}$ is the wavevector. The electron momentum and energy are given by the following expressions:

$$\mathbf{p} = \hbar\mathbf{k}, \qquad \mathcal{E} = \frac{\hbar^2 k^2}{2m_0} \tag{1.22}$$

where m_0 is the free electron mass. It is possible to demonstrate that the solutions of the Schrödinger equation for a periodic potential are *Bloch-Floquet* (or simply Bloch) *functions* so that (*Bloch's theorem*):

$$\psi_{n\mathbf{k}}(\mathbf{r}) = u_{n\mathbf{k}}(\mathbf{r})\, e^{i\mathbf{k}\cdot\mathbf{r}} \tag{1.23}$$

where $u_{n\mathbf{k}}(\mathbf{r})$ has the same periodicity of the crystal:

$$u_{n\mathbf{k}}(\mathbf{r}+\mathbf{R}) = u_{n\mathbf{k}}(\mathbf{r}) \tag{1.24}$$

$\mathbf{k}$ is the wave vector and n refers to the band (as it will be discussed in this chapter). The Bloch functions are the product of a plane wave and a periodic function $u_{n\mathbf{k}}(\mathbf{r})$, with the lattice periodicity. It is evident that the Bloch wavefunction is not periodic, indeed:

$$\psi_{n\mathbf{k}}(\mathbf{r}+\mathbf{R}) = \psi_{n\mathbf{k}}(\mathbf{r})\, e^{i\mathbf{k}\cdot\mathbf{R}} \tag{1.25}$$

while the electron probability density is periodic:

$$|\psi_{n\mathbf{k}}(\mathbf{r}+\mathbf{R})|^2 = |\psi_{n\mathbf{k}}(\mathbf{r})|^2 \tag{1.26}$$

1.3.1 Intuitive proof of the Bloch's theorem in the case of a linear lattice Let us consider a linear lattice with lattice constant a. $U(x)$ is the periodic potential, such that $U(x+ma) = U(x)$, where m is an integer. We assume that the electron distribution has the same periodicity of the lattice, so that:

$$|\psi(x+ma)|^2 = |\psi(x)|^2 \tag{1.27}$$

Equation 1.27 implies that

$$\psi(x+ma) = C\psi(x) \tag{1.28}$$

where C is a quantity satisfying the condition $|C|^2 = 1$. Therefore we may write $C = e^{ikma}$, where k is an arbitrary parameter. From Eq. 1.28 we obtain:

$$\psi(x) = e^{-ikma}\psi(x+ma). \tag{1.29}$$

Multiplying both sides of the previous equation by e^{-ikx}, we obtain

$$\psi(x)e^{-ikx} = e^{-ik(x+ma)}\psi(x+ma). \tag{1.30}$$

thus showing that the function $u(x) = e^{-ikx}\psi(x)$ is periodic with period a. Finally, by writing $\psi(x) = e^{ikx}u(x)$, we obtain the expression of the Bloch's theorem in the case of a linear lattice. ∎

If $\mathbf{G}$ is a translation vector in the reciprocal space, the Bloch function can be written as:

$$\psi_{n\mathbf{k}}(\mathbf{r}) = u_{n\mathbf{k}}(\mathbf{r})\, e^{i\mathbf{k}\cdot\mathbf{r}} = [u_{n\mathbf{k}}(\mathbf{r})e^{-i\mathbf{G}\cdot\mathbf{r}}]e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}} \tag{1.31}$$

which is still a Bloch function since the function between the square brackets in the previous equation has the same periodicity of the lattice (see Eq. 1.11). Therefore the wavefunction 1.31 is the solution of the Schrödinger equation 1.20 for the wavevector $\mathbf{k}+\mathbf{G}$. We conclude that, for a given band n, the eigenstates, $\psi_{n\mathbf{k}}(\mathbf{r})$, and eigenvalues, $\mathcal{E}_n(\mathbf{k})$, are periodic function of $\mathbf{k}$ in the reciprocal lattice:

$$\psi_{n\mathbf{k}}(\mathbf{r}) = \psi_{n,\mathbf{k}+\mathbf{G}}(\mathbf{r}) \tag{1.32}$$

$$\mathcal{E}_n(\mathbf{k}) = \mathcal{E}_n(\mathbf{k}+\mathbf{G}). \tag{1.33}$$

These two important expressions clearly demonstrate that we can restrict our analysis to the first Brillouin zone. In particular, it is possible to represent the dispersion relation $\mathcal{E}(\mathbf{k})$ only in this zone by performing a translation of the outer branches into the first Brillouin zone by the reciprocal lattice wavevector $\mathbf{G}$, so that $\mathbf{k}+\mathbf{G}$ is contained in the first zone. This procedure is called band folding.

We note that the vector $\mathbf{k}$ used in the previous equations plays the same role of the wavevector $\mathbf{k}$ in the case of a free electron. However, there is an important difference: while for free electrons the momentum can be written as $\mathbf{p} = \hbar\mathbf{k}$, in the case of electrons in a periodic structure $\hbar\mathbf{k}$ is not the electron momentum. Indeed, it can be easily demonstrated that the Bloch wavefunction $\psi_{n\mathbf{k}}$ is not an eigenstate of the momentum operator $\mathbf{p} = -i\hbar\nabla$. In fact if we apply the momentum operator to a Bloch function we obtain:

$$-i\hbar\nabla\psi_{n\mathbf{k}} = -i\hbar\nabla\left(u_{n\mathbf{k}}(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{r}}\right) = \hbar\mathbf{k}\psi_{n\mathbf{k}} - i\hbar e^{i\mathbf{k}\cdot\mathbf{r}}\nabla u_{n\mathbf{k}}(\mathbf{r}). \tag{1.34}$$

In contrast, in the case of a free electron, the application of the momentum operator leads to the following equation:

$$-i\hbar\nabla\psi_{\mathbf{k}} = \hbar\mathbf{k}\psi_{\mathbf{k}}, \tag{1.35}$$

thus showing that the wave function $\psi_{\mathbf{k}}$ is an eigenfunction of the momentum operator with the eigenvalue $\hbar\mathbf{k}$. The velocity of the electron with wavevector $\mathbf{k}$ is therefore given by $\mathbf{v} = \hbar\mathbf{k}/m$.

The physical meaning of the term $\hbar\mathbf{k}$ associated to an electron in a crystal and called *crystal momentum* can be understood by considering the equation of motion of an electron in a periodic structure, under the action of an externally applied electromagnetic field. To study the motion of an electron in a particular band, from the quantum-mechanical point of view we have to consider the wave-packet formed by wavefunctions characterized by wavenumbers $\mathbf{k}'$ distributed around a particular $\mathbf{k}$ value:

$$\begin{aligned}\psi_{n\mathbf{k}}(\mathbf{r},t) &= \sum_{\mathbf{k}'} g(\mathbf{k}')\, u_{n\mathbf{k}'}(\mathbf{r}) e^{i(\mathbf{k}'\cdot\mathbf{r}-\omega_n t)} = \\ &= \sum_{\mathbf{k}'} g(\mathbf{k}')\, \psi_{n\mathbf{k}'}(\mathbf{r}) \exp\left(-i\frac{\mathcal{E}_n(\mathbf{k}')}{\hbar}t\right),\end{aligned} \tag{1.36}$$

where we have written $\omega_n = \mathcal{E}_n/\hbar$. It is assumed that the $\mathbf{k}'$ values are within an interval around $\mathbf{k}$ much smaller than the extension of the Brillouin zone. The group velocity of the wave-packet is:

$$\mathbf{v}_{gn} = \frac{\partial\omega_n}{\partial\mathbf{k}} = \frac{1}{\hbar}\frac{\partial\mathcal{E}_n(\mathbf{k})}{\partial\mathbf{k}} \tag{1.37}$$

which can also be written as:

$$\mathbf{v}_{gn} = \frac{1}{\hbar}\nabla_{\mathbf{k}}\mathcal{E}_n(\mathbf{k}) \tag{1.38}$$

The dispersion relation $\mathcal{E}_n(\mathbf{k})$ contains all the effects of the crystal on the motion of the electron. If $\mathbf{E}$ is the applied electric field, the work done by $\mathbf{E}$ on the electron (charge $-e$) during the time interval dt is:

$$d\mathcal{E} = -e\mathbf{E}\cdot d\mathbf{r} = -e\mathbf{E}\cdot\mathbf{v}_g dt = -e\mathbf{E}\cdot\frac{1}{\hbar}\frac{d\mathcal{E}}{d\mathbf{k}}dt. \tag{1.39}$$

Therefore we obtain:

$$\frac{d(\hbar\mathbf{k})}{dt} = -e\mathbf{E} = \mathbf{F}_{ext} \tag{1.40}$$

where $\mathbf{F}_{ext} = -e\mathbf{E}$ is the external force acting on the electron. Upon considering also the effects of the magnetic field, Eq. 1.40 can be written as:

$$\frac{d(\hbar\mathbf{k})}{dt} = -e[\mathbf{E}(\mathbf{r},t) + \mathbf{v}_g(\mathbf{k})\times\mathbf{B}(\mathbf{r},t)]. \tag{1.41}$$

While in the case of free electrons $-e\mathbf{E}$ (or $-e(\mathbf{E}+\mathbf{v}_g\times\mathbf{B})$) is the total force acting on the charged particles (and therefore from Eqs. 1.40 and 1.41 we obtain that $\hbar\mathbf{k}$ is the electron momentum), for an electron in a crystal $-e\mathbf{E}$ represents only the external force, while the total force is given by $\mathbf{F}_{ext} + \mathbf{F}_{int}$, where $\mathbf{F}_{int}$ represents the resultant internal force from the crystal lattice. Therefore:

$$\frac{d\mathbf{p}}{dt} = \mathbf{F}_{ext} + \mathbf{F}_{int}. \tag{1.42}$$

1.4 The concept of effective mass

In the case of free electrons, the energy is given by Eq. 1.22, so that the electron mass can be written as:

$$\frac{1}{m_0} = \frac{1}{\hbar^2}\frac{d^2\mathcal{E}}{dk^2}. \tag{1.43}$$

This expression can be generalized for an arbitrary dispersion relation $\mathcal{E}(\mathbf{k})$. If we differentiate Eq. 1.37 we obtain:

$$\frac{dv_g}{dt} = \frac{1}{\hbar}\frac{\partial^2\mathcal{E}}{\partial t \partial k} = \frac{1}{\hbar}\frac{\partial^2\mathcal{E}}{\partial k^2}\frac{\partial k}{\partial t}. \tag{1.44}$$

By using Eq. 1.40 we have:

$$\frac{dv_g}{dt} = \left(\frac{1}{\hbar^2}\frac{\partial^2\mathcal{E}}{\partial k^2}\right)F_{ext}, \qquad \text{or} \qquad F_{ext} = \frac{\hbar^2}{\partial^2\mathcal{E}/\partial k^2}\frac{dv_g}{dt}. \tag{1.45}$$

Comparing Eq. 1.45 with the second Newton's law we obtain the definition of the *effective mass* m^*:

$$\frac{1}{m^*} = \frac{1}{\hbar^2}\frac{\partial^2\mathcal{E}}{\partial k^2}. \tag{1.46}$$

The previous equation can be generalized for arbitrary dispersion relations to take into account an anisotropic energy surface (this is the case of important semiconductors such as Si and Ge). Since the effective mass is generally dependent on direction, it is a tensor. For a band with anisotropy we may write:

$$\frac{1}{m^*_{ij}} = \frac{1}{\hbar^2}\frac{\partial^2\mathcal{E}}{\partial k_i \partial k_j}, \qquad i,j = x,y,z. \tag{1.47}$$

In order to calculate the dispersion relation $\mathcal{E}_n(\mathbf{k})$, which gives the relationship among electron energy in a crystal, the wavevectors $\mathbf{k}$ in the first Brillouin zone and the band index n, one has to solve the Schrödinger equation obtained by introducing the Bloch wavefunctions 1.23 into Eq. 1.20. The set of curves $\mathcal{E}_n(\mathbf{k})$ defines the band structure of the material. Note that the Bloch wavefunctions are labelled also with respect to the band index n since, for a given value of $\mathbf{k}$, there are many solutions of the Schrödinger equation. Indeed, if we introduce the Bloch function in the Schrödinger equation, we obtain:

$$\left[\frac{\hbar^2}{2m}(-i\nabla + \mathbf{k})^2 + U(\mathbf{r})\right]u_{\mathbf{k}}(\mathbf{r}) = \mathcal{E}_k u_{\mathbf{k}}(\mathbf{r}), \tag{1.48}$$

therefore $u_{\mathbf{k}}(\mathbf{r})$ is determined by the eigenvalue problem

$$H_k u_{\mathbf{k}}(\mathbf{r}) = \mathcal{E}_k u_{\mathbf{k}}(\mathbf{r}), \tag{1.49}$$

with boundary conditions:

$$u_{\mathbf{k}}(\mathbf{r}) = u_{\mathbf{k}}(\mathbf{r} + \mathbf{R}). \tag{1.50}$$

Due to the periodic boundary conditions, this eigenvalue problem is restricted to a single primitive cell of the crystal, so that, in general, an infinite set of solutions exists with discretely spaced eigenvalues, which can be labelled by the band index n.

1.5 Energy bands

It is well know that in atoms, bound electrons occupy discrete energy levels; in solids such discrete levels broaden to form allowed bands, which can be separated by band gaps, as shown in Fig. 1.12. Indeed, when the distance between two initially isolated

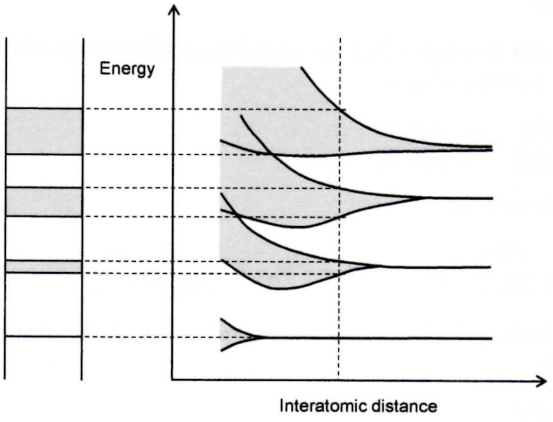

Figure 1.12: Formation of energy bands.

atoms is decreased, their electronic orbitals overlap and the final wavefunction is described by two combinations, $\psi_1 \pm \psi_2$, of the two atomic wavefunctions, which correspond to the bonding and anti-bonding states of the final diatomic molecule. In the case of N atoms, N different orbitals are generated from each orbital of the isolated atom. In a given crystal lattice there are several energy bands, which correspond to the atomic energy levels of the atoms composing the crystal, as schematically shown in Fig. 1.12, and the width of the band is proportional to the strength of the overlap interaction between neighbouring atoms. The energy bands which derive from inner atomic shells are completely filled by electrons, which are more or less localized. In general, many important physical properties of the crystal are determined by electrons contained in bands corresponding to uppermost atomic shells, occupied by the valence electrons. If the uppermost band is not completely filled by electrons at 0 K it is called conduction band. If it is completely filled at 0 K, it is called valence band and the empty band just above it is called conduction band. The material in which the uppermost band is not completely filled is a metal, while in the second case we have an insulator or a semiconductor, depending on the width of the energy gap, $\mathcal{E}_g$, between the valence and the conduction band. As a rule of thumb we have a semiconductor when $0 < \mathcal{E}_g \leq 4$ eV and an insulator when $\mathcal{E} > 4$ eV. This discrimination is not so sharp, indeed diamond, with an energy gap $\mathcal{E} = 5.5$ eV is still considered a semiconductor. The materials with $\mathcal{E} = 0$ are called semi-metals. When a band is completely filled by electrons, such electrons cannot conduct any current.

1.5.1 Electrons and holes in a semiconductor

In the case of a semiconductor, upon increasing the temperature some electrons can be excited from the valence to the conduction band. In this way the valence band is left with a few empty states. Empty states in a band are generally called holes. When

an external electromagnetic field is applied to a semiconductor a hole behaves as a particle with a positive charge e. Let us consider a completely filled band. In this case the sum of all the wave vectors of the electrons is zero:

$$\sum \mathbf{k} = 0 \tag{1.51}$$

the sum being over all states of the considered band in the first Brillouin zone. This is a consequence of the inversion symmetry of the first Brillouin zone, which follows from the symmetry under the inversion $\mathbf{r} \to -\mathbf{r}$ about any lattice point of the crystal lattice. As a result there is the same number of filled states with positive wavenumber $\mathbf{k}$ and with negative wavenumber $-\mathbf{k}$. If an electron with wavenumber $\mathbf{k}_e$ is promoted from the valence to the conduction band, the total wavevector is now $\sum \mathbf{k} = -\mathbf{k}_e$. This wavevector is assigned to the hole, so that the wavevector of the hole, $\mathbf{k}_h$, is given by:

$$\mathbf{k}_h = -\mathbf{k}_e. \tag{1.52}$$

The equation of motion of the missing electron in an electromagnetic field is given by Eq. 1.41:

$$\frac{d(\hbar \mathbf{k}_e)}{dt} = -e(\mathbf{E} + \mathbf{v}_g \times \mathbf{B}), \tag{1.53}$$

therefore the corresponding equation of motion of the hole (with $\mathbf{k}_h = -\mathbf{k}_e$) is:

$$\frac{d(\hbar \mathbf{k}_h)}{dt} = e(\mathbf{E} + \mathbf{v}_g \times \mathbf{B}), \tag{1.54}$$

which is the equation of motion of a particle with positive charge e.

1.6 Calculation of the band structure

Various techniques can be employed to calculate the dependence of energy vs wavevector, i.e., the dispersion relation $\mathcal{E}(\mathbf{k})$, which gives the band structure of the material. It is not the aim of this section to describe in detail these band structure calculations, which are extensively treated in many textbooks. Each method must solve two main problems: (i) the determination of the periodic one-electron potential, $U(\mathbf{r})$, of the crystal, which is contained in the Schrödinger equation; (ii) find the eigenvalues and eigenfunctions of this equation. The various methods differ in the procedures used to solve these two problems. In the following the tight-binding method will be briefly discussed, mainly for didactic reasons, since, in particular situations, it allows one to derive analytic expressions for the dispersion relation $\mathcal{E}(\mathbf{k})$, which are very useful to get a simple physical understanding of the band structure. We will also mention the so-called $\mathbf{k} \cdot \mathbf{p}$ method. Many other methods can be employed: for example the pseudopotential method, the cellular (Wigner-Seitz) method; the so-called muffin-tin methods, which include the augmented plane wave (APW) method, and others.

1.6.1 Tight-binding method

The tight-binding method (TBM) is an empirical approach since experimental inputs have to be used to fit the actual band structure. In the previous section we have seen how energy bands are created upon progressively reducing the distances between

the atoms composing the crystal. In the final configuration the atomic wavefunctions of neighbouring atoms can be more or less overlapped: the TBM can be properly applied when the overlap of the atomic wavefunctions cannot be neglected but it is not too large, so that the atomic description can still be used. In this case the atomic wavefunctions can be used as a basis set for the Bloch functions. Strictly speaking, in order to obtain the eigenstates of the valence electrons of a crystal, all atomic wavefunctions of the atoms composing the crystal are required, since only all atomic orbitals compose a complete basis set. However, the largest contribution is given by the atomic orbitals of the valence shells of the free atoms: within the TBM only these orbitals are considered. For example, for the elemental semiconductors of group IV of the periodic table, the valence shell orbitals are formed by the $2s$ and $2p$ states in the case of C, by the $3s$ and $3p$ states for Si and by the $4s$ and $4p$ states for Ge. In the case of semiconductors formed by different elements, the valence shell orbitals of the various atoms have to be considered. For example, in the case of GaAs the $4s$ state and the $4p$ states of Ga and the $4s$ and $4p$ states of As must be considered. In the above examples the valence shells of the atoms are formed by $s-$type or $p-$type orbitals.

We will assume a lattice with a single atom basis. We will further assume that the bound states of the atomic Hamiltonian, H_{at}, are well localized so that the wavefunction $\psi_n(\mathbf{r})$ for an atom placed at the origin ($\mathbf{r} = 0$) becomes very small at a distance from the origin of the order of the lattice constant. $\psi_n(\mathbf{r})$ is given by

$$H_{at}\psi_n = \mathcal{E}_n\psi_n \tag{1.55}$$

We assume that close to each lattice point the crystal Hamiltonian, H, can be approximated by H_{at}. The crystal Hamiltonian can be written as:

$$H = H_{at} + \Delta U(\mathbf{r}) \tag{1.56}$$

where $\Delta U(\mathbf{r})$ is the correction to the atomic potential, which has to be added to obtain the periodic potential of the crystal, as shown schematically in Fig. 1.13. If $\Delta U(\mathbf{r})$ vanishes in the region where the atomic orbital $\psi_n(\mathbf{r})$ is localized ($\mathbf{r} = 0$ in Fig. 1.13), $\psi_n(\mathbf{r})$ will be also a solution of the Schrödinger equation with the Hamiltonian given by Eq. 1.56. In this case the crystal wavefunction $\psi_{\mathbf{k}}(\mathbf{r})$ can be written as a linear combination of atomic wavefunctions. Since $\psi(\mathbf{r})$ must satisfy the Bloch condition:

$$\psi_{\mathbf{k}}(\mathbf{r} + \mathbf{R}_0) = e^{i\mathbf{k}\cdot\mathbf{R}_0}\psi_{\mathbf{k}}(\mathbf{r}) \tag{1.57}$$

the linear combination can be written as follows:

$$\psi_{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}}\psi_n(\mathbf{r} - \mathbf{R}) \tag{1.58}$$

where the wavevector $\mathbf{k}$ varies within the first Brillouin zone and the sum over $\mathbf{R}$ spans all the crystal lattice. Such wavefunction satisfies the requirement of a Bloch wavefunction, indeed, considering a generic lattice vector $\mathbf{R}_0$:

$$\begin{aligned} \psi_{\mathbf{k}}(\mathbf{r} + \mathbf{R}_0) &= \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}}\psi_n(\mathbf{r} - \mathbf{R} + \mathbf{R}_0) = \\ &= e^{i\mathbf{k}\cdot\mathbf{R}_0}\sum_{\mathbf{R}} e^{i\mathbf{k}\cdot(\mathbf{R}-\mathbf{R}_0)}\psi_n[\mathbf{r} - (\mathbf{R} - \mathbf{R}_0)] \end{aligned} \tag{1.59}$$

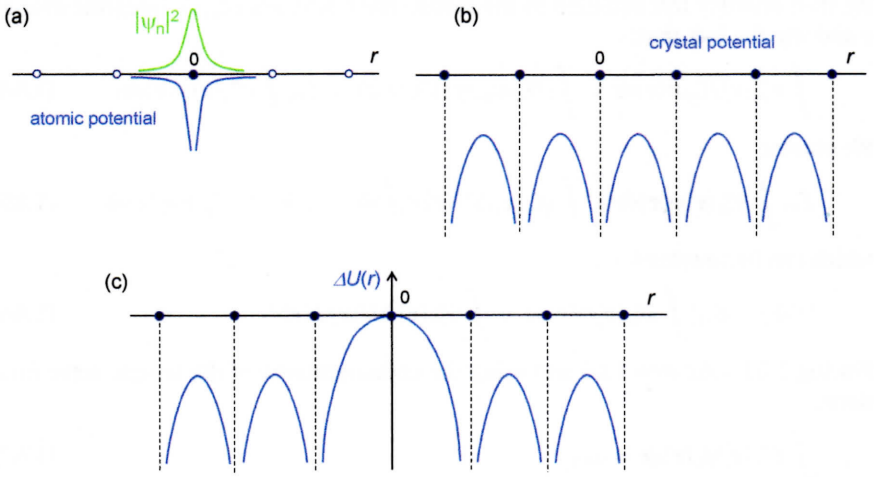

Figure 1.13: (a) Atomic potential of a single atom (blue curve), the green curve schematically represents the squared modulus of an atomic wavefunction, $|\psi_n|^2$, localized at the origin ($r = 0$). (b) Periodic crystal potential. (c) $\Delta U(\mathbf{r})$: correction to the atomic potential to obtain the periodic crystal potential.

$\mathbf{R} - \mathbf{R}_0$ is simply another crystal translation vector and, since the sum over $\mathbf{R}$ spans all the crystal, we can replace it by another translation vector, $\mathbf{R}'$, so that

$$\psi_{\mathbf{k}}(\mathbf{r} + \mathbf{R}_0) = e^{i\mathbf{k}\cdot\mathbf{R}_0} \sum_{\mathbf{R}'} e^{i\mathbf{k}\cdot\mathbf{R}'} \psi_n(\mathbf{r} - \mathbf{R}') = e^{i\mathbf{k}\cdot\mathbf{R}_0} \psi_{\mathbf{k}}(\mathbf{r}) \tag{1.60}$$

It is evident that the energy $\mathcal{E}(\mathbf{k})$ associated to the wavefunction 1.58 is simply the energy of the atomic level, $\mathcal{E}_n$. This is related to our initial assumption of a $\Delta U(\mathbf{r})$ term which vanishes where $\psi_n(\mathbf{r})$ is localized: this assumption is not realistic. Instead, we have to assume that $\Delta U(\mathbf{r})$ is small but not precisely zero where $\psi_n(\mathbf{r})$ becomes small, so that the product $\Delta U(\mathbf{r})\psi_n(\mathbf{r})$ is very small but not zero. Therefore, a more realistic solution for the crystal Schrödinger equation, still satisfying the Bloch conditions, can be written as:

$$\psi(\mathbf{r}) = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} \phi(\mathbf{r} - \mathbf{R}) \tag{1.61}$$

where $\phi(\mathbf{r})$ is very close, but not exactly equal, to the atomic wavefunction $\psi_n(\mathbf{r})$. We can assume that $\phi(\mathbf{r})$ can be written as a linear combination of a small number of localized atomic wavefunctions:

$$\phi(\mathbf{r}) = \sum_n b_n \psi_n(\mathbf{r}) \tag{1.62}$$

for this reason the TBM is also known as the *linear combination of atomic orbitals (LCAO) method.*
The Schrödinger equation is now:

$$H\psi(\mathbf{r}) = [H_{at} + \Delta U(\mathbf{r})]\psi(\mathbf{r}) = \mathcal{E}(\mathbf{k})\psi(\mathbf{r}). \tag{1.63}$$

We then multiply this equation by the atomic wave function $\psi_m^*(\mathbf{r})$, integrate over all $\mathbf{r}$ and use the fact that:

$$\int \psi_m^*(\mathbf{r}) H_{at} \psi(\mathbf{r}) d\mathbf{r} = \int (H_{at}\psi_m(\mathbf{r}))^* \psi(\mathbf{r}) d\mathbf{r} = \mathcal{E}_m \int \psi_m^*(\mathbf{r})\psi(\mathbf{r}) d\mathbf{r}. \tag{1.64}$$

We obtain:

$$\mathcal{E}_m \int \psi_m^*(\mathbf{r})\psi(\mathbf{r})d\mathbf{r} + \int \psi_m^*(\mathbf{r})\Delta U(\mathbf{r})\psi(\mathbf{r})d\mathbf{r} = \mathcal{E}(\mathbf{k}) \int \psi_m^*(\mathbf{r})\psi(\mathbf{r})d\mathbf{r} \tag{1.65}$$

which can be rewritten as:

$$[\mathcal{E}(\mathbf{k}) - \mathcal{E}_m] \int \psi_m^*(\mathbf{r})\psi(\mathbf{r})d\mathbf{r} = \int \psi_m^*(\mathbf{r})\Delta U(\mathbf{r})\psi(\mathbf{r})d\mathbf{r} \tag{1.66}$$

Placing 1.61-1.62 into 1.66 and using the orthonormality of the atomic wave functions:

$$\int \psi_m^*(\mathbf{r})\psi_n(\mathbf{r})d\mathbf{r} = \delta_{mn} \tag{1.67}$$

we can write an eigenvalue equation from which it is possible to obtain the coefficients b_n and the Bloch energies $\mathcal{E}(\mathbf{k})$. In the sum over $\mathbf{R}$ we separate the terms with $\mathbf{R} = 0$ and $\mathbf{R} \neq 0$

$$\begin{aligned}
[\mathcal{E}(\mathbf{k}) - \mathcal{E}_m]b_m &= -[\mathcal{E}(\mathbf{k}) - \mathcal{E}_m] \sum_n \Big(\sum_{\mathbf{R}\neq 0} \int \psi_m^*(\mathbf{r})\psi_n(\mathbf{r}-\mathbf{R}) e^{i\mathbf{k}\cdot\mathbf{R}} d\mathbf{r} \Big) b_n + \\
&+ \sum_n \Big(\int \psi_m^*(\mathbf{r})\Delta U(\mathbf{r})\psi_n(\mathbf{r})d\mathbf{r} \Big) b_n + \\
&+ \sum_n \Big(\sum_{\mathbf{R}\neq 0} \int \psi_m^*(\mathbf{r})\Delta U(\mathbf{r})\psi_n(\mathbf{r}-\mathbf{R}) e^{i\mathbf{k}\cdot\mathbf{R}} d\mathbf{r} \Big) b_n
\end{aligned} \tag{1.68}$$

In a semiconductor we have to consider the valence shell of the atoms from which the crystal is formed, therefore we need to consider an $s-$level and triply degenerate $p-$levels. For this reason Eq. 1.68 gives rise to a set of four homogeneous equations, whose eigenvalues give the dispersion relations $\mathcal{E}(\mathbf{k})$ for the $s-$band and the three $p-$bands and whose solutions $b(\mathbf{k})$ give the amplitudes of the linear combination of atomic orbitals (see Eq. 1.62) forming the wavefunction $\phi(\mathbf{r})$ in the first Brillouin zone. In this case we have to solve a 4×4 secular problem. The calculations are rather complex. To get some physical insight into the problem we will solve a simple problem of one atom basis with only an $s-$function.

1.6.2 Crystal with one-atom basis and single atomic orbital

Since we have to consider only one atomic level, the coefficients $\{b_m\}$ are zero except for the $s-$level, where $b_s = 1$. A single equation results in this case. If $\mathcal{E}_s$ is the energy of the atomic s-level and we write $\psi_s(\mathbf{r}) = \phi(\mathbf{r})$, the following functions can be introduced:

$$\alpha(\mathbf{R}) = \int \phi^*(\mathbf{r})\phi(\mathbf{r}-\mathbf{R})d\mathbf{r} \tag{1.69}$$

$$\beta = -\int \phi^*(\mathbf{r})\Delta U(\mathbf{r})\phi(\mathbf{r})d\mathbf{r} = -\int \Delta U(\mathbf{r})|\phi(\mathbf{r})|^2 d\mathbf{r} \tag{1.70}$$

$$\gamma(\mathbf{R}) = -\int \phi^*(\mathbf{r})\Delta U(\mathbf{r})\phi(\mathbf{r}-\mathbf{R})d\mathbf{r} \tag{1.71}$$

Equation 1.68 can be rewritten in terms of the previous quantities:

$$\mathcal{E}(\mathbf{k}) = \mathcal{E}_s - \frac{\beta + \sum \gamma(\mathbf{R}) e^{i\mathbf{k}\cdot\mathbf{R}}}{1 + \sum \alpha(\mathbf{R}) e^{i\mathbf{k}\cdot\mathbf{R}}} \tag{1.72}$$

Since we are considering an s-level, $\alpha(-\mathbf{R}) = \alpha(\mathbf{R})$, therefore given the fact that $\Delta U(-\mathbf{r}) = \Delta U(\mathbf{r})$, we also have that $\gamma(-\mathbf{R}) = \gamma(\mathbf{R})$. In Eq. 1.72 we can ignore the sum $\sum \alpha(\mathbf{R}) e^{i\mathbf{k}\cdot\mathbf{R}}$ with respect to 1 since $\phi(\mathbf{r})$ is well localized so that the overlapping integral 1.69, which contains the product of two atomic wavefunctions centered on different sites, is very small. Moreover, we can make another simplifying assumption if we assume that in the sum $\sum \gamma(\mathbf{R}) e^{i\mathbf{k}\cdot\mathbf{R}}$ we have to consider only the nearest neighbours (n.n.). Equation 1.72 can be written in a very simple form:

$$\mathcal{E}(\mathbf{k}) = \mathcal{E}_s - \beta - \sum_{\text{n.n.}} \gamma(\mathbf{R}) e^{i\mathbf{k}\cdot\mathbf{R}} \tag{1.73}$$

In the case of real semiconductors we cannot limit ourself to consider a crystal with one-atom basis and one atomic orbital. We already observed that the outermost electrons are placed in 4 atomic orbitals (one s and three p orbitals, p_x, p_y and p_z). Moreover, the corresponding lattice has a two-atom basis. In this case the crystal states can be expanded as linear combinations of two Bloch states, corresponding to the two atoms of the basis. Therefore, by using Eqs. 1.61 and 1.62 we can write:

$$\psi(\mathbf{r}) = \sum_{\mathbf{R}} \sum_{m=1}^{4} \sum_{j=1}^{2} C_{mj}(\mathbf{k}) \psi_{mj}(\mathbf{r} - \mathbf{r}_j - \mathbf{R}) e^{i\mathbf{k}\cdot\mathbf{R}} \tag{1.74}$$

where the index m refers to the different atomic orbitals (s, p_x, p_y and p_z), the index j refers to the atom in the basis, $\mathbf{r}_j$ gives the position of the $j-$th atom in the unit cell. For example, in the case of the zincblende lattice (which is the structure of several semiconductors) for each atom of the two-atom basis we have an s orbital and three p orbitals, thus leading to a 8×8 secular problem (8 homogeneous equations of the form 1.68). Considering also the spin degeneracy we would end up with a 16×16 secular problem, thus increasing the complexity of the tight-binding calculations.

In the following section we will consider two very simple examples of lattices with one-atom basis and a single atomic state ($s-$state) in order to see two simple applications of the TBM. These simple applications will allow us to discuss a few important physical properties, which can be generalized to more realistic situations.

1.6.3 Linear lattice

In the case of a *linear lattice* with lattice constant a the tight-binding wavefunction 1.61 can be written as:

$$\psi(x) = \sum_{m} e^{imka} \phi(x - ma). \tag{1.75}$$

From Eq. 1.73 we can immediately obtain the dispersion relation $\mathcal{E}(k)$ (which is shown in the first Brillouin zone in Fig. 1.14):

$$\mathcal{E}(k) = \mathcal{E}_s - \beta - \gamma(e^{ika} + e^{-ika}) = \mathcal{E}_s - \beta - 2\gamma\cos(ka) \tag{1.76}$$

The width of the band is 4γ and it is proportional to the overlap integral. This is reasonable, since an increase of the overlap integral γ leads to a stronger interaction,

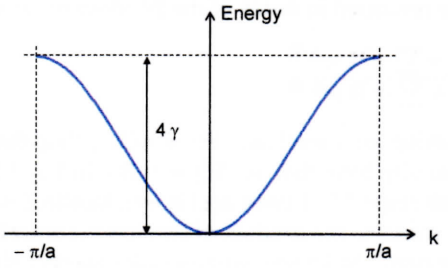

Figure 1.14: Dispersion relation, $\mathcal{E}(k)$, in the first Brillouin zone of a linear lattice.

and consequently a wider bandwidth. When the electron is near the bottom of the band, where $ka \ll 1$, we have:

$$\mathcal{E}(k) \simeq \mathcal{E}_s - \beta - 2\gamma\left(1 - \frac{1}{2}k^2a^2\right) = \mathcal{E}_s - \beta - 2\gamma + \gamma k^2a^2 \tag{1.77}$$

which has the same parabolic dependence on k as the dispersion relation of a free electron. Therefore we may conclude that an electron in that region of the $k-$space behaves as a free electron with an effective mass:

$$m^* = \frac{\hbar^2}{2\gamma a^2} \tag{1.78}$$

which shows that the effective mass is inversely proportional to the overlap integral γ. This is intuitively reasonable, since if the overlap between atomic wavefunctions increases, it is easier for an electron to hop from one atomic site to another, thus leading to a reduction of its effective mass (less inertia for the electronic motion in the crystal).

Upon considering the other atomic energy levels, we end up with various energy

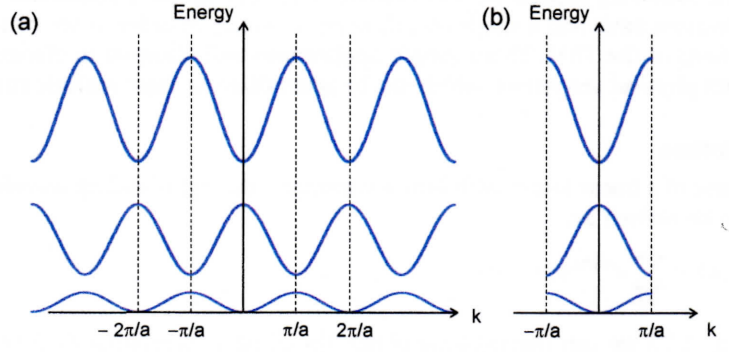

Figure 1.15: Dispersion relation, $\mathcal{E}(k)$, in a linear lattice calculated by using the TBM.

bands, as shown in Fig. 1.15(a). Figure 1.15(b) shows the resulting reduced zone scheme, we have discussed as a consequence of Eq. 1.33 (band-folding procedure). Upon using Eq. 1.77 it is possible to calculate the group velocity of the electron and

the effective mass in the first Brillouin zone (not only ariund $k = 0$):

$$v_g = \frac{1}{\hbar}\frac{\partial \mathcal{E}}{\partial k} = \frac{2\gamma a}{\hbar}\sin(ka), \tag{1.79}$$

$$m^* = \frac{\hbar^2}{\partial^2 \mathcal{E}/\partial k^2} = \frac{\hbar^2}{2\gamma a^2 \cos(ka)}. \tag{1.80}$$

Figure 1.16 shows v_g and m^* as a function of k. Note that in the central region of the

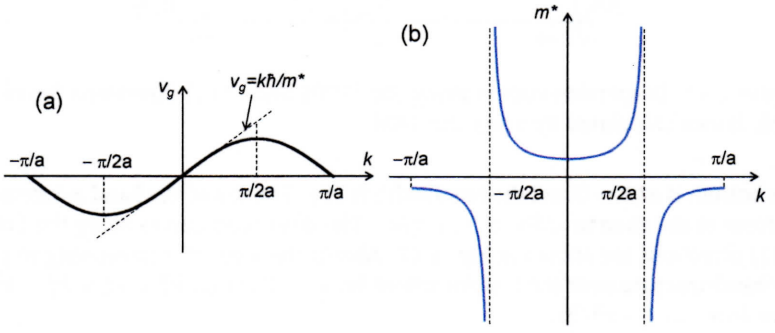

Figure 1.16: (a) Group velocity and (b) effective mass in the first Brillouin zone of a linear lattice.

first Brillouin zone the group velocity is close to the free-electron velocity, $\hbar k/m^*$. In this region the electron can move almost freely through the lattice. This is no more true approaching the borders of the Brillouin zone, $k = \pm\pi/a$: here the perturbation of the lattice is strong. This can be easily understood in terms of Bragg reflection in the linear lattice. Indeed, the Bragg condition for constructive interference of the waves scattered from each lattice point is $k = n\pi/a$. In this situation we have a superposition between a forward ($k = \pi/a$) and a backward ($k = -\pi/a$) wave thus leading to a standing wave, which does not propagate in the crystal ($v_g = 0$).

1.6.4 Simple cubic lattice

Let us solve the problem for a simple cubic (sc) lattice. The six nearest neighbours are at:

$$a(\pm 1, 0, 0)$$
$$a(0, \pm 1, 0)$$
$$a(0, 0, \pm 1)$$

Equation 1.73 can be written as:

$$\begin{aligned}\mathcal{E}(\mathbf{k}) &= \mathcal{E}_s - \beta - \gamma(e^{ik_x a} + e^{-ik_x a} + e^{ik_y a} + e^{-ik_y a} + e^{ik_z a} + e^{-ik_z a}) \\ &= \mathcal{E}_s - \beta - 2\gamma(\cos k_x a + \cos k_y a + \cos k_z a)\end{aligned} \tag{1.81}$$

Along the [100] direction we have ($k_x = k, k_y = k_z = 0$):

$$\mathcal{E}(\mathbf{k}) = \mathcal{E}_s - \beta - 2\gamma(\cos ka + 2) \tag{1.82}$$

The bottom of the band is at $k = 0$ and the bandwidth is 4γ. Along the [111] direction ($k_x^2 = k_y^2 = k_z^2 = k^2/3$) we have:

$$\mathcal{E}(\mathbf{k}) = \mathcal{E}_s - \beta - 6\gamma \cos\frac{ka}{\sqrt{3}} \tag{1.83}$$

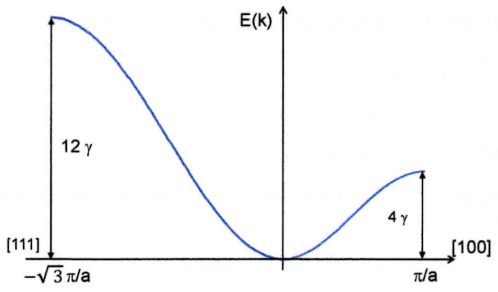

Figure 1.17: Dispersion curves along the [100] and [111] directions for an simple cubic lattice calculated by using the TBM.

The bottom is at $k = 0$ and the bandwidth is 12γ. The top of the band is located at the extreme of the zone at $\sqrt{3}[\pi/a, \pi/a, \pi/a]$. The dispersion curves along the [100] and [111] directions are shown in Fig. 1.17. Also in this case, it is interesting to examine the bandstructure near the Γ point where $ka \ll 1$. Putting: $k_x^2 = k_y^2 = k_z^2 = k^2/3$, we have ($\cos x \simeq 1 - x^2/2$):

$$\begin{aligned} \mathcal{E}(\mathbf{k}) &= \mathcal{E}_s - \beta - 6\gamma \cos\frac{ka}{\sqrt{3}} \simeq \mathcal{E}_s - \beta - 6\gamma\left(1 - \frac{k^2a^2}{6}\right) \\ &= \mathcal{E}_s - \beta - 6\gamma + \gamma k^2 a^2 \end{aligned} \tag{1.84}$$

Comparing with the free electron problem solution, as in the previous example, we can obtain a simple expression for the effective mass (valid when $ka \ll 1$):

$$m^* = \frac{\hbar^2}{2\gamma a^2} \tag{1.85}$$

1.6.5 Band structure of semiconductors calculated by TBM

The band structure for GaAs calculated by using TBM is shown in Fig. 1.18. The top of the valence band and the bottom of the conduction band (referred to as the bandedges of the semiconductor) are both at the Γ−point: GaAs is therefore a direct gap semiconductor. The states near the bottom of the conduction band have the symmetry of the s−state, whereas the states at the top of the valence band have the symmetry of the p−states. Generally, the states in the valence band can be written as linear combinations of p−orbitals. We note that the Bloch-lattice function maintains most of the symmetry properties as the original atomic orbitals. The bandgap of the semiconductor and other properties in the high-symmetry points can be well fitted by using experimental inputs. Nevertheless, the calculated crystal structure shown in Fig. 1.18 do not reproduce the real structure of GaAs. This is clearly evident looking at the top of the valence band: Fig. 1.18 shows a 3-fold degeneracy, resulting from the degeneracy of the p_x, p_y and p_z states. In contrast, measurements show that the top of the valence band is 2-fold degenerate and a third band is present below the valence bandedge. This discrepancy is due to the fact that an important parameter has not been included so far in our discussion: the spin. The first result is to double the degeneracy of each state, both in the valence and conduction band, without changing the dispersion curves. The second crucial effect is the spin-orbit interaction, which leads to changes of the energy curves in the valence band. Spin-orbit interaction is a relativistic effect: when an electron moves about the nucleus at

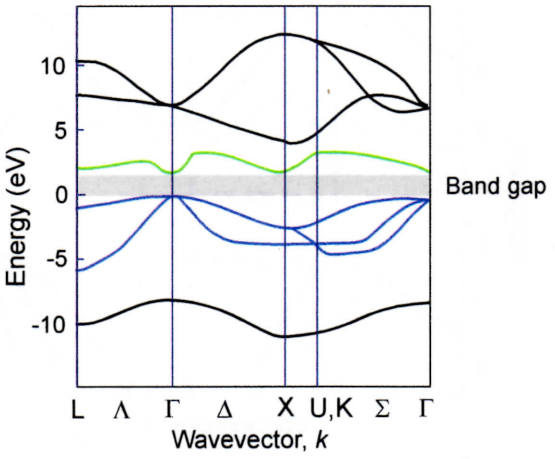

Figure 1.18: Calculated tight-binding band-structure of GaAs without the effects of the spin-orbit coupling.

relativistic velocity, the electric field generated by the nucleus Lorentz-transforms to a magnetic field, which interacts with the electron spin magnetic moment. In atoms the spin-orbit interaction splits each electron energy level with a nonzero orbital quantum number ℓ. Spin-orbit splitting in the bandstructure of crystals is due to the same effect. Spin-orbit interaction leads to modifications in the valence band, whose states are primarily $p-$states (with orbital angular momentum $\hbar$). Without entering into any detail, spin-orbit interaction can be included in the tight-binding calculation by adding an additional term, H_{SO}, in the Hamiltonian:

$$H_{SO} = \xi \, \mathbf{L} \cdot \mathbf{S} \tag{1.86}$$

where $\mathbf{L}$ is the operator for orbital angular momentum, $\mathbf{S}$ is the operator for spin angular momentum and ξ can be treated as a constant. Upon considering the spin-orbit interaction, the dispersion curves calculated for GaAs modify as shown in Fig. 1.19. The top of the valence band loses part of its degeneracy: a *split-off band* is formed, separated by an energy offset Δ_{SO} from the top of the valence band. The spin-orbit splittings for various semiconductors are reported in Table 1.2.

The spin-orbit splitting energy Δ_{SO} of semiconductors increases as the fourth power

Semiconductor	Δ_{SO} (eV)
Si	0.044
Ge	0.29
GaAs	0.34
InAs	0.41
InSb	0.85

Table 1.2: Spin-orbit splitting energy for various semiconductors.

of the atomic number of the constituent elements. Intuitively this can be understood as follows. If the atomic number Z (which gives the number of protons in the nucleus) increases, also the electric field seen by the valence electrons increases and

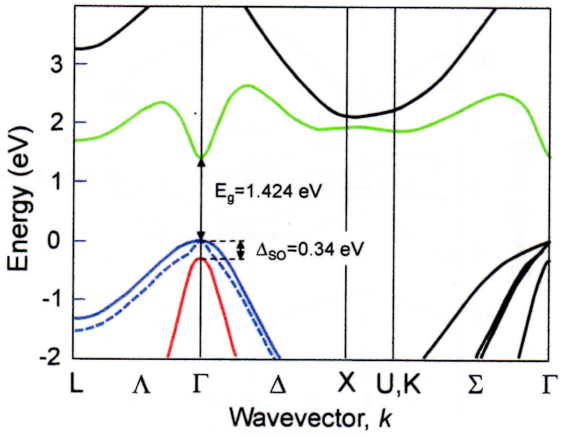

Figure 1.19: Calculated tight-binding band-structure of GaAs including the effects of the spin-orbit coupling. The split-off band is represented by the red curve. Δ_{SO} is the spin-orbit splitting energy.

consequently also the interaction between the orbital momentum and the spin. Δ_{SO} is small for materials composed by light atoms, as Si, and may be quite large in comparison with the energy gap in semiconductors composed by heavy atoms, as InSb. Note that the conduction band is not affected by spin-orbit interaction (due to the $s-$symmetry) and remains doubly degenerate.

Figure 1.20 shows schematically the dispersion curves near the bandedges of a semiconductor with direct or indirect (dashed conduction band curve) bandgap. The doubly degenerate valence bands (4-fold degeneracy with spin) at the $\Gamma-$point are characterized by two different curvatures: these bands are called *heavy-hole* (HH) and *light-hole* (LH) bands. The band with the smaller curvature is the HH band, since the effective mass is inversely proportional to the band curvature.

So far we have considered only direct bandgap semiconductors such as GaAs; there are semiconductors, like Si and Ge, where the $k-$value of the bandedge of the conduction band is different from the $k-$value corresponding to the top of the valence band: these semiconductors are called *indirect*. For example, in Si the conduction bandedge is close to the $X-$point, in Ge the conduction bandedge is at the $L-$point, while for both semiconductors the valence bandedge is at the $\Gamma-$point. In the case of indirect semiconductors the states at the conduction bandedge present a strong anisotropy and can be described by combinations of $s-$ and $p-$type states.

As it will be clear in the next chapters, for most practical situations, only the dispersion relations close to the top of the valence band and to the bottom of the conduction bands have to be considered. For this reason in most cases parabolic dispersion relations can be considered. In this case we can write:

$$\mathcal{E}(k) = \mathcal{E}_g + \frac{\hbar^2 k^2}{2m_c} \tag{1.87}$$

in the conduction band, where $\mathcal{E}_g$ is the energy gap and m_c is the effective mass in the conduction band, and

$$\mathcal{E}(k) = -\frac{\hbar^2 k^2}{2m_{lh}}, \qquad \mathcal{E}(k) = -\frac{\hbar^2 k^2}{2m_{hh}} \tag{1.88}$$

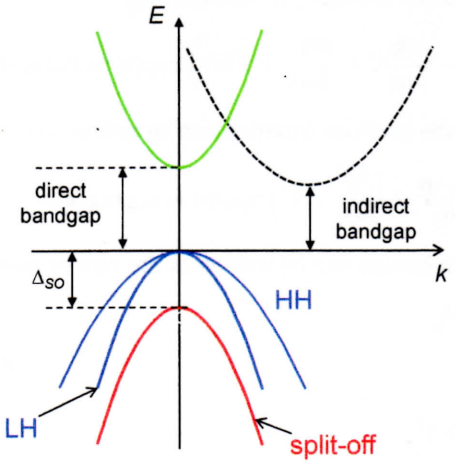

Figure 1.20: Schematic representation of the dispersion curves near the bandedges of a semiconductor with direct or indirect (dashed conduction band curve) bandgap. HH: heavy-hole valence band; LH: light-hole valence band.

in the valence band, where m_{lh} and m_{hh} are the effective masses for the light-hole and for the heavy-hole, respectively. Table 1.3 reports the values of the energy gap and effective masses for a few typical III-V semiconductors.

Material	$\mathcal{E}_g$ (eV) (0 K)	$\mathcal{E}_g$ (eV) (300 K)	m_c/m_0	m_{lh}/m_0	m_{hh}/m_0
GaAs	1.52	1.424	0.067	0.082	0.45
InP	1.42	1.27	0.08	0.089	0.6
InAs	0.43	0.354	0.023	0.025	0.41
InSb	0.23	0.17	0.014	0.016	0.40

Table 1.3: Energy gap (in eV) and effective masses (in unit of the free-electron mass m_0) of typical III-V semiconductors.

1.7 The $\mathbf{k} \cdot \mathbf{p}$ method

In this section we will briefly discuss another method, called $\mathbf{k} \cdot \mathbf{p}$ method, frequently used to calculate the band structure of semiconductors, which is quite accurate near the bandedges. Here we will just illustrate the basic idea at the heart of this method. In the following we will use a few concepts and mathematical procedures which will be discusses in more details in Chapter 3.
According to the Bloch theorem, the solution of the one-electron Schrödinger equation can be written as:

$$\psi_{n\mathbf{k}} = e^{i\mathbf{k}\cdot\mathbf{r}} u_{n\mathbf{k}}(\mathbf{r}) \tag{1.89}$$

where n is the band index, $\mathbf{k}$ is the wavector in the first Brillouin zone and $u_{n\mathbf{k}}(\mathbf{r})$ has the periodicity of the crystal. When 1.89 is substituted in the Schrödinger equation

the following equation in $u_{n\mathbf{k}}(\mathbf{r})$ is obtained:

$$\left(-\frac{\hbar^2}{2m_0}\nabla^2 - i\frac{\hbar\mathbf{k}}{m_0}\nabla + \frac{\hbar^2k^2}{2m_0} + V(\mathbf{r})\right)u_{n\mathbf{k}}(\mathbf{r}) = \mathcal{E}_{n\mathbf{k}}u_{n\mathbf{k}}(\mathbf{r}). \tag{1.90}$$

Since $\mathbf{p} = -i\hbar\nabla$, the previous equation can be written as:

$$\left(\frac{p^2}{2m_0} + \frac{\hbar\mathbf{k}\cdot\mathbf{p}}{m_0} + \frac{\hbar^2k^2}{2m_0} + V\right)u_{n\mathbf{k}}(\mathbf{r}) = \mathcal{E}_{n\mathbf{k}}u_{n\mathbf{k}}(\mathbf{r}). \tag{1.91}$$

Therefore the Hamiltonian can be written as the sum of two terms:

$$H = H_0 + H'_{\mathbf{k}} \tag{1.92}$$

where

$$H_0 = \frac{p^2}{2m_0} + V \tag{1.93}$$

is the unperturbed Hamiltonian and

$$H'_{\mathbf{k}} = \frac{\hbar\mathbf{k}\cdot\mathbf{p}}{m_0} + \frac{\hbar^2k^2}{2m_0} \tag{1.94}$$

is the perturbation Hamiltonian. In the $\Gamma-$point ($\mathbf{k} = 0$) Eq. 1.91 gives:

$$\left(\frac{p^2}{2m_0} + V\right)u_{n0} = H_0u_{n0} = \mathcal{E}_{n0}u_{n0} \qquad n = 1, 2, 3, \ldots \tag{1.95}$$

The solutions of Eq. 1.95 form a complete and orthonormal set of basis functions. Now we can use the terms $\hbar^2\mathbf{k}\cdot\mathbf{p}/m_0$ and $\hbar^2k^2/2m_0$ of the Hamiltonian as perturbations of the first and second order in $\mathbf{k}$. For this reason the $\mathbf{k}\cdot\mathbf{p}$ model is quite accurate for small values of k. We note that the method can be used to calculate $\mathcal{E}_n(\mathbf{k})$ around any value $\mathbf{k}_0$ by expanding Eq. 1.91 around $\mathbf{k}_0$ if the wavefunctions and the energies at $\mathbf{k}_0$ are known. Now, if we assume that the band structure presents a stationary point (either a maximum or a minimum) at the energy $\mathcal{E}_{n0}$ and that the band is nondegenerate at this energy, by using standard perturbation theory it is possible to show that the wavefunctions $u_{n\mathbf{k}}$ and the energies $\mathcal{E}_{n\mathbf{k}}$ can be written as:

$$u_{n\mathbf{k}} = u_{n0} + \frac{\hbar}{m_0}\sum_{n'\neq n}\frac{\langle u_{n0}|\mathbf{k}\cdot\mathbf{p}|u_{n'0}\rangle}{\mathcal{E}_{n0} - \mathcal{E}_{n'0}}u_{n'0} \tag{1.96}$$

$$\mathcal{E}_{n\mathbf{k}} = \mathcal{E}_{n0} + \frac{\hbar^2k^2}{2m_0} + \frac{\hbar^2}{m_0^2}\sum_{n'\neq n}\frac{|\langle u_{n0}|\mathbf{k}\cdot\mathbf{p}|u_{n'0}\rangle|^2}{\mathcal{E}_{n0} - \mathcal{E}_{n'0}} \tag{1.97}$$

Note that the linear terms in k are not present since we are assuming that $\mathcal{E}_{n0}$ represents a maximum or a minimum of the band structure. In 1.96 and 1.97 the matrix elements $\langle u_{n0}|\mathbf{k}\cdot\mathbf{p}|u_{n'0}\rangle$ can be written as:

$$\langle u_{n0}|\mathbf{k}\cdot\mathbf{p}|u_{n'0}\rangle = \int u^*_{n0}\mathbf{k}\cdot\mathbf{p}u_{n'0}\,d\mathbf{r}. \tag{1.98}$$

Since $\mathbf{k}$ is a vector of real numbers, these matrix elements can be rewritten as:

$$\langle u_{n0}|\mathbf{k}\cdot\mathbf{p}|u_{n'0}\rangle = \mathbf{k}\cdot\langle u_{n0}|\mathbf{p}|u_{n'0}\rangle. \tag{1.99}$$

For small values of k it is very useful to write $\mathcal{E}_{n\mathbf{k}}$ as:

$$\mathcal{E}_{n\mathbf{k}} = \mathcal{E}_{n\mathbf{0}} + \frac{\hbar^2 k^2}{2m^*} \tag{1.100}$$

where m^* is the effective mass. By comparing Eqs. 1.100 and 1.97 we obtain the following expression for the effective mass:

$$\frac{1}{m^*} = \frac{1}{m_0} + \frac{2}{m_0^2 k^2} \sum_{n' \neq n} \frac{|\langle u_{n\mathbf{0}}|\mathbf{k}\cdot\mathbf{p}|u_{n'\mathbf{0}}\rangle|^2}{\mathcal{E}_{n\mathbf{0}} - \mathcal{E}_{n'\mathbf{0}}}. \tag{1.101}$$

This result is quite instructive since it states that the effective mass is not equal to the mass m_0 of the free electron as a result of the coupling between electronic states in different bands via the term $\mathbf{k}\cdot\mathbf{p}$. Moreover the electron effective mass decreases upon decreasing the width of the gap $\mathcal{E}_{n\mathbf{0}} - \mathcal{E}_{n'\mathbf{0}}$ if the dominant term in 1.101 is given by the coupling between the conduction and valence band.

As an example, from Eq. 1.97 it is possible to obtain an approximated expression of the dispersion curve for the conduction band assuming that $\mathcal{E}_{n\mathbf{0}}$ corresponds to the bottom of the conduction band, $\mathcal{E}_{c\mathbf{0}}$. The sum over n' can include only the terms with energies close to the top of the valence band, where the energy difference at the denominator, $\mathcal{E}_{n\mathbf{0}} - \mathcal{E}_{n'\mathbf{0}}$ is smallest, since these terms give the most important contribution. Indeed, the relative importance of a band n' to the effective mass is controlled by the energy gap between the two bands. Therefore we can reasonably assume:

$$\mathcal{E}_{n\mathbf{0}} - \mathcal{E}_{n'\mathbf{0}} = \mathcal{E}_{c\mathbf{0}} - \mathcal{E}_{v\mathbf{0}} = \mathcal{E}_g. \tag{1.102}$$

From Eq. 1.99 we have:

$$|\langle u_{n\mathbf{0}}|\mathbf{k}\cdot\mathbf{p}|u_{n'\mathbf{0}}\rangle|^2 = |\mathbf{k}\cdot\langle u_{c\mathbf{0}}|\mathbf{p}|u_{v\mathbf{0}}\rangle|^2 = k^2|p_{cv}|^2 \tag{1.103}$$

where p_{cv} is the momentum matrix element between the two bandedge Bloch functions:

$$p_{cv} = \langle u_{c\mathbf{0}}|\mathbf{p}|u_{v\mathbf{0}}\rangle = \int u^*_{c\mathbf{0}}(\mathbf{r})\mathbf{p}u^*_{v\mathbf{0}}(\mathbf{r})\, d\mathbf{r} \tag{1.104}$$

Assuming the zero of the energy axis at the top of the dispersion curve, from Eq. 1.97 we can write:

$$\mathcal{E}_c(\mathbf{k}) = \mathcal{E}_{c\mathbf{0}} + \frac{\hbar^2 k^2}{2m_0} + \frac{\hbar^2 k^2}{m_0^2}\frac{|p_{cv}|^2}{\mathcal{E}_g} = \mathcal{E}_g + \frac{\hbar^2 k^2}{2m_0} + \frac{\hbar^2 k^2}{m_0^2}\frac{|p_{cv}|^2}{\mathcal{E}_g} \tag{1.105}$$

and from Eq. 1.101 we have:

$$\frac{1}{m^*} = \frac{1}{m_0} + \frac{2}{m_0^2}\frac{|p_{cv}|^2}{\mathcal{E}_g} \tag{1.106}$$

A usually used parameter is the *Kane energy*:

$$\mathcal{E}_P = \frac{2}{m_0}|p_{cv}|^2, \tag{1.107}$$

so that the dispersion curve 1.105 can be written as:

$$\mathcal{E}_c(\mathbf{k}) = \mathcal{E}_g + \frac{\hbar^2 k^2}{2m_0}\left(1 + \frac{\mathcal{E}_P}{\mathcal{E}_g}\right) \tag{1.108}$$

and the effective mass is

$$\frac{1}{m^*} = \frac{1}{m_0}\left(1 + \frac{\mathcal{E}_P}{\mathcal{E}_g}\right). \tag{1.109}$$

The Kane energy is usually much larger than the energy gap and it is, to a good approximation, almost the same for most III-V semiconductors. Table 1.4 reports the Kane energy values of a few semiconductors. For the conduction band of direct gap semiconductors, if the band gap decreases also the electron effective mass decreases. In other words, the bigger the energy gap, the smaller is the effect on the effective mass. For example, in InSb the conduction band effective mass is very small, $m^* = m_0/77$ as a consequence of the very small energy gap $\mathcal{E}_g = 250$ meV.

	$\mathcal{E}_P$ (eV)
GaAs	28.8
AlAs	21.1
InAs	21.5
InP	20.7
GaN	25.0
AlN	27.1

Table 1.4: Kane energy values of a few semiconductors (from: I. Vurgaftman *et al.*, J. Appl. Phys. **89**, 5815-5875 (2001)).

1.8 Bandstructures of a few semiconductors

In this final section we will briefly discuss relevant properties of the bandstructure of a few representative semiconductors. We will concentrate mainly on the characteristics of the bandedges since most of the optical properties of the semiconductors are determined by regions of the bandstructure close to the top of the valence band and to the bottom of the conduction band. We will consider three semiconductors: Si, GaAs and GaN.

1.8.1 Silicon

We have already mentioned that Si crystal has a diamond structure, with the first Brillouin zone in the form of a truncated octahedron, as shown in Fig. 1.11(b). Silicon is an indirect semiconductor, for this reason it is not extensively used in optical devices. The top of the valence band is at the $\Gamma-$point, while the bottom of the conduction band is at the point $\frac{2\pi}{a}(0.85, 0, 0)$, close to the $X-$point (the direction [100] from the $\Gamma-$point to the $X-$point is called $\Delta-$line). Since there are six equivalent $\Delta-$lines, there are six equivalent minima in the conduction band at $k = \frac{2\pi}{a}(\pm 0.85, 0, 0)$, $\frac{2\pi}{a}(0, \pm 0.85, 0)$ and $\frac{2\pi}{a}(0, 0 \pm 0.85)$. The corresponding energy gap is 1.1 eV. The effective mass at these anisotropic minima is characterized by a longitudinal mass along the six equivalent [100] directions and two transverse masses in the plane orthogonal to the longitudinal direction. The longitudinal mass is $m_l = 0.98\ m_0$, while the transverse mass is $m_t = 0.19\ m_0$. The constant energy surfaces near the conduction band minima in Si can be written as:

$$\mathcal{E}(\mathbf{k}) = \frac{\hbar^2 k_x^2}{2m_l} + \frac{\hbar^2(k_y^2 + k_z^2)}{2m_t} \tag{1.110}$$

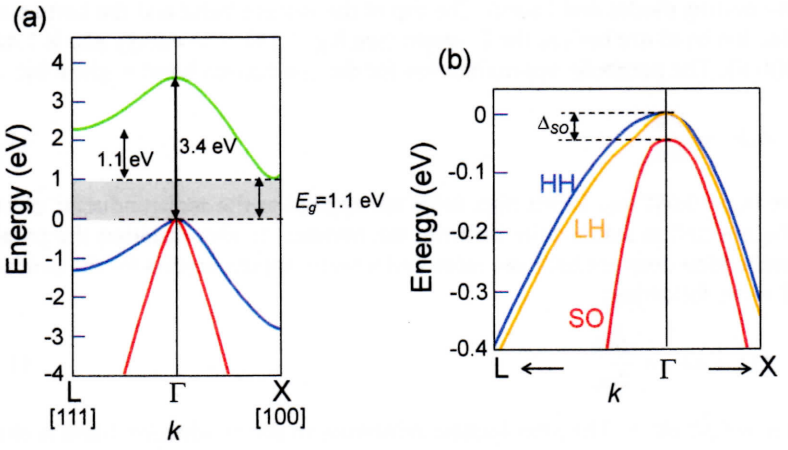

Figure 1.21: (a) Bandstructure of Si. (b) Zoom of a region of the valence band around the Γ−point, showing the heavy-hole (HH), the light-hole (LH) and split-off (SO) bands.

and are ellipsoids, as shown in Fig. 1.22. Equation 1.110 represents the parabolic band approximation. The direct energy gap is $\sim$ 3.4 eV. There is also a second minimum in the conduction band at the point L, which is located about 1.1 eV above the bandedge. The top of the valence band at the Γ−point present the HH and LH degeneracy. The top of the split-off band is just 44 meV below the top of the valence band, as shown in Fig. 1.21(b). The light-hole mass is $m_{lh} = 0.16\, m_0$, the heavy-hole mass is $m_{hh} = 0.46\, m_0$, the split-off hole mass is $m_{h,SO} = 0.29\, m_0$.

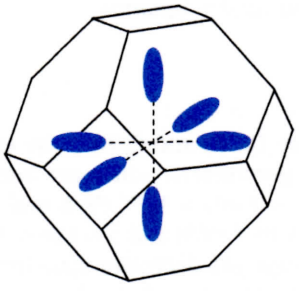

Figure 1.22: Constant energy surfaces near the conduction band minima in Si. There are six equivalent valley in Si at the bandedge.

1.8.2 Gallium Arsenide

Gallium arsenide is a III-V semiconductor. The electron configuration of isolated atoms is [Ar]$3d^{10}4s^24p^1$ for Ga and [Ar]$3d^{10}4s^24p^3$ for As. When forming a GaAs crystal the outermost shell of each atom is hybridized in the form sp^3 so that each atom is surrounded by four neighbour atoms at the apex of a regular tetrahedron. GaAs is characterized by a zincblende crystal structure with a lattice constant $a = 5.65$ Å. GaAs is a direct bandgap semiconductor and it is largely used for optical devices

(light-emitting diodes and lasers). The top of the valence band and the bottom of the conduction band are both at the $\Gamma-$point (see Fig. 1.19). The energy gap is 1.424 eV (at 300 K). The parabolic approximation for the conduction band is given by:

$$\mathcal{E}(k) = \frac{\hbar^2 k^2}{2m_c} \tag{1.111}$$

where $m_c = 0.067\ m_0$. When high fields are applied to the semiconductor, electrons can be injected in states quite far from the minima. In this situation the parabolic approximation does not hold any more and a better approximation for the conduction band is the following:

$$\mathcal{E}(1 + \alpha\mathcal{E}) = \frac{\hbar^2 k}{2m_c} \tag{1.112}$$

with $\alpha = 0.67\ \text{eV}^{-1}$. The next highest minimum in the conduction band is close to the $L-$point. The separation between the Γ and L minima is 0.29 eV. The effective mass at the L minimum is much larger than the corresponding mass at the $\Gamma-$point: $m^* \approx 0.25\ m_0$. The next higher energy minimum in the conduction band is at the $X-$point; the separation between the Γ and X minima is 0.48 eV. The corresponding electron effective mass is $m_X^* \approx 0.6\ m_0$. When the semiconductor is subject to high electric fields, both the L and X valleys can be populated by electrons, in addition to the Γ valley. Therefore these regions can be quite important to describe the electronic processes in GaAs. The valence band is characterized by the light- and heavy-hole bands and by the split-off band. The light- and heavy-hole masses are $m_{lh} = 0.082\ m_0$ and $m_{hh} = 0.45\ m_0$, respectively. The split-off energy is quite large: $\Delta = 0.34$ eV, so that, in general, this band is always completely filled by electrons and does not play any relevant role in the electronic and optical properties of GaAs.
The temperature dependence of the energy gap can be well approximated by the following expression, called *Varshni equation*:

$$\mathcal{E}_g = \mathcal{E}_g(0) - \frac{\alpha T^2}{T + \beta} \tag{1.113}$$

where T is the temperature (in K) and $\mathcal{E}_g(0)$ is the energy gap at $T = 0$. $\mathcal{E}_g(0)$, α and β are suitable fitting parameters, which are material-specific. In the case of GaAs $\mathcal{E}_g(0) = 1.519$ eV, $\alpha = 5.405 \times 10^{-4}$ eV/K and $\beta = 204$ K. In the case of Si $\mathcal{E}_g(0) = 1.166$ eV, $\alpha = 4.73 \times 10^{-4}$ eV/K and $\beta = 636$ K. This is a quite general behavior in semiconductors: the energy gap decreases upon increasing the temperature. It is possible to show that the temperature dependence of the energy gap is due to the electron-phonon interaction in the semiconductor, which depends on the amplitude of the phonons and on the corresponding coupling constants.

1.8.3 Gallium Nitride

Gallium nitride is another III-V semiconductor with direct bandgap, largely used in optoelectronic devices like light-emitting diodes and lasers. It has a large direct energy gap $\mathcal{E}_g = 3.45$ eV, whose dependence on temperature can be taken into account by using the Varshni equation with $\mathcal{E}_g(0) = 3.51$ eV, $\alpha = 9.09 \times 10^{-4}$ eV/K and $\beta = 830$ K (in the range 293 K $< T <$ 1237 K). It usually presents the hexagonal wurtzite crystal structure. Also in this semiconductor each atom presents sp^3 hybridization and forms tetrahedral bonds with the nearest-neighbours, with a bonding angle of 109.47° and a

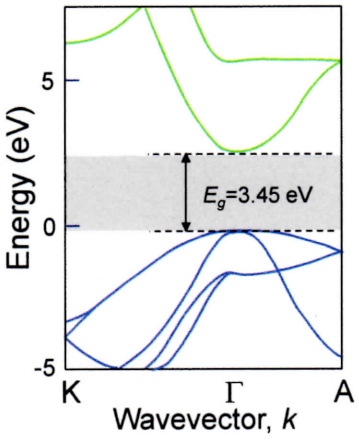

Figure 1.23: Bandstructure of GaN.

bond length of 19.5 nm. The large energy gap is directly related to the large bonding strength in GaN. The first Brillouin zone of a hexagonal crystal is a hexagonal prism with height $2\pi/c$, as shown in Fig. 1.11(c). The spin-orbit splitting energy is $\Delta = 15.5$ meV. The band structure of GaN near the bandedges is shown in Fig. 1.23. The electron effective mass in the conduction band ($\Gamma-$minimum) is $m_c = 0.20\ m_0$; in the valence band the HH effective mass along the axis a is $m_{hh,a} = 1.65\ m_0$, along the $c-$axis $m_{hh,c} = 1.1\ m_0$, the LH masses are: $m_{lh,a} = 0.15\ m_0$ and $m_{lh,c} = 1.1\ m_0$.

2. Electrons in semiconductors

2.1 Introduction

In the previous chapter we have seen that in a semiconductor at $T = 0$ K the valence band is completely filled with electrons while the conduction band is empty. In this way no current can flow in the material. When electrons are promoted in the conduction band, holes are generated in the valence band and current can now flow in the semiconductor. In this chapter we will discuss the main properties of carriers in semiconductors. The density of states will be first calculated and the essential concepts of carrier statistics in semiconductors will be discussed.

2.2 Density of states

We assume that the semiconductor is divided in adjacent boxes with edges $\mathbf{L}_i = N_i \mathbf{a}_i$ ($i = 1, 2, 3$), where $\mathbf{a}_i$ are primitive vectors and N_i can be considered arbitrarily large. We further assume that the crystal repeats itself exactly in the same way from one box to the other, so that:

$$\psi(\mathbf{r} + N_i \mathbf{a}_i) = \psi(\mathbf{r}) \tag{2.1}$$

The previous equation is the mathematical formulation of the so-called *Born-von Karman boundary conditions*. Since $\psi(\mathbf{r})$ is a Bloch function Eq. 2.1 can be written as follows:

$$\psi_{n\mathbf{k}}(\mathbf{r} + N_i \mathbf{a}_i) = e^{i\mathbf{k} \cdot N_i \mathbf{a}_i} \psi_{n\mathbf{k}}(\mathbf{r}) = \psi_{n\mathbf{k}}(\mathbf{r}) \tag{2.2}$$

so that

$$e^{iN_i \mathbf{k} \cdot \mathbf{a}_i} = 1. \tag{2.3}$$

The wavevector $\mathbf{k}$ can be written as:

$$\mathbf{k} = \kappa_1 \mathbf{b}_1 + \kappa_2 \mathbf{b}_2 + \kappa_3 \mathbf{b}_3 \tag{2.4}$$

where $\mathbf{b}_j$ are the primitive vectors in the reciprocal space. Since $\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi\delta_{ij}$, Eq. 2.3 can be written as follows:

$$e^{i2\pi\kappa_i N_i} = 1, \tag{2.5}$$

so that

$$\kappa_i = \frac{m_i}{N_i}, \tag{2.6}$$

where m_i are integers. We conclude that the allowed $\mathbf{k}$ values are given by the following expression:

$$\mathbf{k} = \frac{m_1}{N_1}\mathbf{b}_1 + \frac{m_2}{N_2}\mathbf{b}_2 + \frac{m_3}{N_3}\mathbf{b}_3. \tag{2.7}$$

Therefore each state in the $\mathbf{k}-$space occupies a volume Δk given by the volume of a parallelepiped with edges $\mathbf{b}_i/N_i$, given by:

$$\Delta k = \frac{\mathbf{b}_1}{N_1} \cdot \left(\frac{\mathbf{b}_2}{N_2} \times \frac{\mathbf{b}_3}{N_3}\right) = \frac{1}{N}\mathbf{b}_1 \cdot (\mathbf{b}_2 \times \mathbf{b}_3), \tag{2.8}$$

where $N = N_1N_2N_3$ is the total number of primitive cells in the crystal and $\mathbf{b}_1 \cdot (\mathbf{b}_2 \times \mathbf{b}_3)$ represents the volume of a primitive cell in the reciprocal lattice. It is possible to demonstrate that the reciprocal lattice primitive vectors, $\mathbf{b}_i$, satisfy the following relation:

$$\mathbf{b}_1 \cdot (\mathbf{b}_2 \times \mathbf{b}_3) = \frac{(2\pi)^3}{\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)}. \tag{2.9}$$

Since the quantity at the denominator of the previous expression represents the volume, V_u, of a Bravais lattice primitive cell in the direct lattice, so that $V_u = V/N$, the volume of the reciprocal lattice primitive cell is $(2\pi)^3 N/V$. Therefore Eq. 2.8 can be written as:

$$\Delta k = \frac{(2\pi)^3}{V}. \tag{2.10}$$

Note that in the case of a simple cubic crystal $N_i = L/a$, $\mathbf{a}_i = a\mathbf{u}_i$, $\mathbf{b}_i = (2\pi/a)\mathbf{u}_i$ and Eq. 2.7 gives:

$$\mathbf{k} = m_1\frac{2\pi}{L}\mathbf{u}_x + m_2\frac{2\pi}{L}\mathbf{u}_y + m_3\frac{2\pi}{L}\mathbf{u}_z \tag{2.11}$$

and the volume Δk per allowed k value is given by

$$\Delta k = \frac{(2\pi)^3}{L^3}, \tag{2.12}$$

in agreement with Eq. 2.10.
In the case of a two-dimensional lattice the area per allowed k value is:

$$\Delta k = \frac{(2\pi)^2}{A}, \tag{2.13}$$

where, in the case of a square lattice $A = L^2$. In the case of a one-dimensional semiconductor we have:

$$\Delta k = \frac{2\pi}{L}. \tag{2.14}$$

Exercise 2.1 Demonstrate Eq. 2.9.

We can write $\mathbf{b}_1$ by using Eq.1.6, so that:

$$\mathbf{b}_1 \cdot (\mathbf{b}_2 \times \mathbf{b}_3) = 2\pi \frac{(\mathbf{a}_2 \times \mathbf{a}_3) \cdot (\mathbf{b}_2 \times \mathbf{b}_3)}{\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)} \tag{2.15}$$

The previous expression can be rewritten by using the following identity related to the cross product:

$$(\mathbf{a} \times \mathbf{b}) \cdot (\mathbf{c} \times \mathbf{d}) = (\mathbf{a} \cdot \mathbf{c})(\mathbf{b} \cdot \mathbf{d}) - (\mathbf{a} \cdot \mathbf{d})(\mathbf{b} \cdot \mathbf{c}) \tag{2.16}$$

where **a**, **b**, **c** and **d** are generic vectors. By using Eq. 2.16 in 2.15 one obtains:

$$\mathbf{b}_1 \cdot (\mathbf{b}_2 \times \mathbf{b}_3) = 2\pi \frac{(\mathbf{a}_2 \cdot \mathbf{b}_2)(\mathbf{a}_3 \cdot \mathbf{b}_3) - (\mathbf{a}_2 \cdot \mathbf{b}_3)(\mathbf{a}_3 \cdot \mathbf{b}_2)}{\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)} \tag{2.17}$$

Since $\mathbf{b}_i \cdot \mathbf{a}_j = 2\pi\delta_{ij}$, the previous expression gives:

$$\mathbf{b}_1 \cdot (\mathbf{b}_2 \times \mathbf{b}_3) = \frac{(2\pi)^3}{\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)}. \tag{2.18}$$

■

2.2.1 Bulk (three-dimensional) semiconductors

We can now calculate the density of states, $\rho(\mathcal{E})$, in a semiconductor as a function of energy. The physical meaning of $\rho(\mathcal{E})$ is the following: $\rho(\mathcal{E})\, d\mathcal{E}$ gives the number of states per unit volume with energy in the range between $\mathcal{E}$ and $\mathcal{E} + d\mathcal{E}$. The first step is the calculation of the density of states in the **k**−space, $\rho(k)$. We will assume a parabolic dispersion curve:

$$\mathcal{E} = \frac{\hbar^2 k^2}{2m^*} + \mathcal{E}_0. \tag{2.19}$$

In the **k**−space the states with energy between $\mathcal{E}$ and $\mathcal{E} + d\mathcal{E}$ are the states characterized by k values contained in the region between two spherical surfaces with radii k and $k + dk$, whose volume is $4\pi k^2\, dk$. Since the volume occupied by a single state is $(2\pi)^3/V$, the number of states with k value in the region k and $k + dk$ is:

$$\frac{4\pi k^2 dk}{8\pi^3} V = \frac{k^2}{2\pi^2} V\, dk \tag{2.20}$$

Therefore the density of states in the k−space is given by:

$$\rho(k)\, dk = \frac{k^2}{2\pi^2} dk, \qquad \rho(k) = \frac{k^2}{2\pi^2} \tag{2.21}$$

Upon taking into account the electron spin degeneracy, so that each state can be occupied by two electrons with opposite spin, we obtain the following expression:

$$\rho(k) = \frac{k^2}{\pi^2} \tag{2.22}$$

In order to calculate $\rho(\mathcal{E})$, we have to use the dispersion relation 2.19 and impose the conservation of the number of states in the **k**−space and in the $\mathcal{E}$−space:

$$\rho(\mathcal{E})\, d\mathcal{E} = \rho(k)\, dk \tag{2.23}$$

If we differentiate Eq. 2.19 we obtain:

$$d\mathcal{E} = \frac{\hbar^2 k}{m^*}\, dk \tag{2.24}$$

From Eqs. 2.19, 2.23 and 2.24 we have:

$$\rho(\mathcal{E}) = \frac{1}{2\pi^2}\left(\frac{2m^*}{\hbar^2}\right)^{3/2}\sqrt{\mathcal{E}-\mathcal{E}_0} \tag{2.25}$$

If we refer to the conduction band, in Eq. 2.25 we have $\mathcal{E}_0 = \mathcal{E}_c$, where $\mathcal{E}_c$ is the bandedge (bottom energy of the conduction band).
We note here, without entering into details, that in direct bandgap semiconductors m^* is simply given by the effective mass for the conduction band: $m^* = m_c$. In the case of indirect semiconductors m^* in the conduction band is an average effective mass of the conduction band, given by:

$$m^* = (g^2 m_1 m_2 m_3)^{1/3}, \tag{2.26}$$

where m_1, m_2 and m_3 are the effective masses along the three principal axis and g is the degeneracy factor. For example, for the $X-$valley of silicon $g = 6$ and Eq. 2.26 gives:

$$m_c^* = 6^{2/3}(m_l m_t^2)^{1/3}. \tag{2.27}$$

In the same way, the density of states in the valence band is given by:

$$\rho_v(\mathcal{E}) = \frac{1}{2\pi^2}\left(\frac{2m_v}{\hbar^2}\right)^{3/2}\sqrt{\mathcal{E}_v-\mathcal{E}} \tag{2.28}$$

where $\mathcal{E}_v$ is the bandedge of the valence band (i.e., the valence band maximum) and m_v is given by:

$$m_v = (m_{hh}^{3/2} + m_{lh}^{3/2})^{2/3} \tag{2.29}$$

Figure 2.1(a) shows the density of states in the conduction band.
Note that in III-V semiconductors $m_c \ll m_v$, therefore the density of states in the conduction band is typically smaller than the density of states in the valence band. Moreover, since $m_{lh} \ll m_{hh}$ the density of states of light-holes is a small fraction of the density of states of the heavy-hole states. Therefore light-hole states can be typically neglected in comparison with heavy-holes (this is true in the case of bulk semiconductors, not in the case of quantum-well semiconductors).

2.2.2 Two- and one-dimensional semiconductors

It is also simple to calculate the density of states for a two-dimensional semiconductor (still assuming a parabolic dispersion relation). In this case the states with energy in the range between $\mathcal{E}$ and $\mathcal{E} + d\mathcal{E}$, in the $\mathbf{k}-$space are represented by states with k values contained in the region between two circles with radii k and $k + dk$, with area $2\pi k\, dk$. Since the area occupied by each state is $(2\pi)^2/L^2$ and considering the spin degeneracy we have:

$$\rho(k)\, dk = 2\frac{2\pi k\, dk}{4\pi^2}\frac{L^2}{L^2} = \frac{k}{\pi}\, dk. \tag{2.30}$$

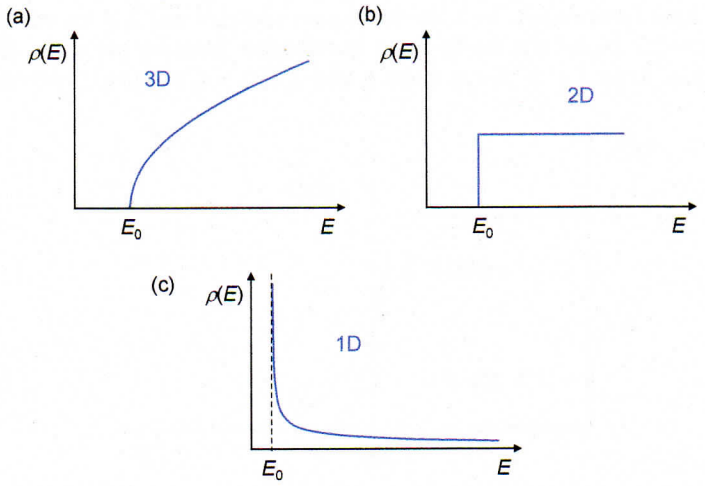

Figure 2.1: Density of states in (a) three-dimensional, (b) two-dimensional and (c) one-dimensional systems.

Since $\rho(\mathcal{E})\, d\mathcal{E} = \rho(k)\, dk$ we obtain:

$$\rho(\mathcal{E}) = \frac{m^*}{\pi\hbar^2}. \tag{2.31}$$

Figure 2.1(b) shows the density of states in the conduction band.
In a one-dimensional system the states with energy between $\mathcal{E}$ and $\mathcal{E} + d\mathcal{E}$ in the $\mathbf{k}-$space are represented by states in two (linear) regions with width dk. Since each state occupies a region with length $2\pi/L$, we obtain:

$$\rho(k) = \frac{2}{\pi} \tag{2.32}$$

and assuming a parabolic dispersion curve the corresponding expression as a function of energy is:

$$\rho_c(\mathcal{E}) = \frac{\sqrt{2m_c}}{\pi\hbar}(\mathcal{E} - \mathcal{E}_c)^{-1/2} \tag{2.33}$$

for the conduction band and

$$\rho_v(\mathcal{E}) = \frac{\sqrt{2m_c}}{\pi\hbar}(\mathcal{E}_v - \mathcal{E})^{-1/2} \tag{2.34}$$

for the valence band. Figure 2.1(c) shows the density of states in the conduction band.

2.3 Carrier statistics in semiconductors

We assume that the semiconductor is in thermal equilibrium. Electrons and holes are fermions, therefore they obey the *Fermi-Dirac statistics*. The probability that a state with energy $\mathcal{E}$ is filled by an electron when the semiconductor is at thermal equilibrium at temperature T is given by the Fermi-Dirac distribution function

$$f(\mathcal{E}) = \frac{1}{1 + \exp(\mathcal{E} - \mathcal{E}_F)/k_B T} \tag{2.35}$$

where $\mathcal{E}_F$ is the *Fermi energy*, k_B is the Boltzmann constant and T is the temperature in K. Figure 2.2 shows the Fermi-Dirac distribution function for three different temperatures. From Eq. 2.35 it is evident that $f(\mathcal{E}_F) = 1/2$. $f(\mathcal{E})$ is very small for

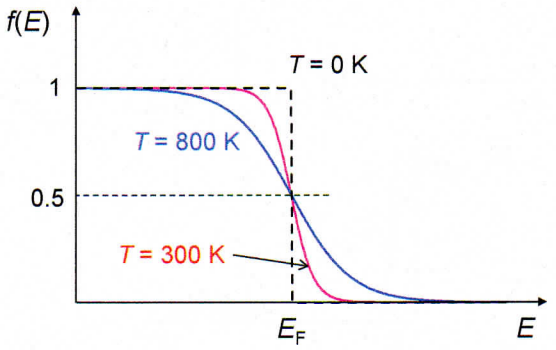

Figure 2.2: Fermi-Dirac distribution function at 0 K (dashed black line), 300 K (red solid line) and 800 K (blue solid line).

energies larger than $\mathcal{E}_F$, meaning that only a few electrons are found in high-energy states at thermal equilibrium. At $T = 0$ the Fermi-Dirac function changes abruptly at $\mathcal{E} = \mathcal{E}_F$ from $f(\mathcal{E}) = 1$ when $\mathcal{E} < \mathcal{E}_F$ to $f(\mathcal{E}) = 0$ when $\mathcal{E} > \mathcal{E}_F$ as shown in Fig. 2.2, so that all allowed states below $\mathcal{E}_F$ are filled with electrons while all states above $\mathcal{E}_F$ are empty.
The probability that a state with energy $\mathcal{E}$ is empty is given by:

$$\overline{f}(\mathcal{E}) = 1 - f(\mathcal{E}) = \frac{1}{1 + \exp(\mathcal{E}_F - \mathcal{E})/k_B T} \tag{2.36}$$

The number of electrons per unit volume in the conduction band is given by:

$$n = \int_{cb} \rho_c(\mathcal{E}) f(\mathcal{E})\, d\mathcal{E} \tag{2.37}$$

where the integral is calculated over the conduction band. In the same way the density of holes in the valence band can be calculated as follows:

$$p = \int_{vb} \rho_v(\mathcal{E}) \overline{f}(\mathcal{E})\, d\mathcal{E} \tag{2.38}$$

By using Eq. 2.25 in Eq. 2.37 one obtains:

$$n = \frac{1}{2\pi^2} \left(\frac{2m_c}{\hbar^2} \right)^{3/2} \int_{\mathcal{E}_c}^{\infty} \sqrt{\mathcal{E} - \mathcal{E}_c} \frac{1}{1 + \exp(\mathcal{E} - \mathcal{E}_F)/k_B T}\, d\mathcal{E} \tag{2.39}$$

If we introduce a new variable x defined as $x = (\mathcal{E} - \mathcal{E}_c)/k_B T$, Eq. 2.39 can be written as follows:

$$n = \frac{1}{4} \left(\frac{2m_c k_B T}{\pi \hbar^2} \right)^{3/2} \frac{2}{\sqrt{\pi}} \int_0^{\infty} \frac{\sqrt{x}}{1 + e^{x-y}}\, dx \tag{2.40}$$

where:

$$y = \frac{\mathcal{E}_F - \mathcal{E}_c}{k_B T} \tag{2.41}$$

Equation 2.40 can be rewritten as follows:

$$n = N_c F_{1/2}(y) = N_c F_{1/2}\left(\frac{\mathcal{E}_F - \mathcal{E}_c}{k_B T}\right), \tag{2.42}$$

where:

$$N_c = \frac{1}{4}\left(\frac{2m_c k_B T}{\pi\hbar^2}\right)^{3/2} \tag{2.43}$$

is called *effective density of states* in the conduction band and $F_{1/2}(y)$ is the *Fermi function* of order $n = 1/2$, shown in Fig. 2.3. In the case of large negative arguments (i.e., $y < 0$ and $|y| \gg 1$) $F_{1/2}(y) \approx \exp(y)$, which is the Boltzmann distribution function. When $y \gg 1$ we have:

$$F_{1/2}(y) \approx \frac{2}{\sqrt{\pi}}\frac{2}{3}y^{3/2}. \tag{2.44}$$

These two approximations are shown by dashed curves in Fig. 2.3. A few values of the Fermi function $F_{1/2}(y)$ are reported in Table 2.1, for y values between -5 and 10.

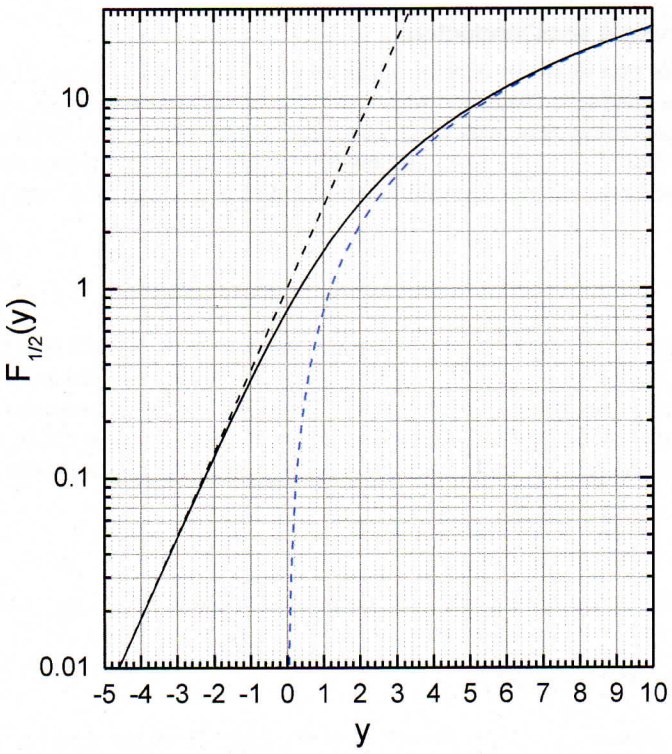

Figure 2.3: Fermi function $F_{1/2}(y)$ (solid black line) and approximations: $\exp(y)$ (black dashed line); $\frac{2}{\sqrt{\pi}}\frac{2}{3}y^{3/2}$ (blue dashed line).

y	$F_{1/2}(y)$
-5	6.722×10^{-3}
-4	1.820×10^{-2}
-3	4.893×10^{-2}
-2	1.293×10^{-1}
-1	3.278×10^{-1}
0	7.652×10^{-1}
1	1.576
2	2.824
3	4.488
4	6.512
5	8.844
6	11.447
7	14.291
8	17.355
9	20.624
10	24.085

Table 2.1: A few values of the Fermi function $F_{1/2}(y)$, for y values between -5 and 10.

2.3.1 Non-degenerate semiconductors

In the case of non-degenerate semiconductors the Fermi energy is well below the bottom of the conduction band and well above the top of the valence band; we assume that the Fermi energy is at least $3k_BT$ away from either band edge. In this case we can assume that $(\mathcal{E} - \mathcal{E}_F) \gg k_BT$, and the Fermi-Dirac distribution function can be approximated by the simpler Maxwell-Boltzmann distribution function (non-quantum regime):

$$f_c(\mathcal{E}) \approx \exp\left(-\frac{\mathcal{E} - \mathcal{E}_F}{k_BT}\right) \tag{2.45}$$

which is valid when we remove the limitations imposed by Pauli exclusion principle. This is reasonable since at energies well above Fermi energy the probability to find an electron in an allowed state is so small that the Pauli exclusion principle does not have practical effects and the classical Maxwell-Boltzmann statistics can be used. As already mentioned, since in this case y is negative and $|y| \gg 1$, one obtains:

$$F_{1/2} \approx e^y \frac{2}{\sqrt{\pi}} \int_0^\infty \frac{\sqrt{x}}{e^x}\, dx = e^y \tag{2.46}$$

so that the electron density in the conduction band is given by (see Eq. 2.42) by:

$$n = N_c \exp\left(-\frac{\mathcal{E}_c - \mathcal{E}_F}{k_BT}\right) \tag{2.47}$$

From this expression it is clear the origin of the name "effective density of states" for N_c. Indeed it is as all the energy states in the conduction band were effectively concentrated into a single energy level, $\mathcal{E}_c$, which can allocate N_c electrons per unit volume. Therefore, the electron density is given by the product of N_c and the probability that the state with energy $\mathcal{E}_c$ is occupied.

In the same way it is possible to calculate the hole density in the valence band:

$$p = N_v \exp\left(-\frac{\mathcal{E}_F - \mathcal{E}_v}{k_B T}\right) \tag{2.48}$$

where:

$$N_v = \frac{1}{4}\left(\frac{2m_v k_B T}{\pi\hbar^2}\right)^{3/2}. \tag{2.49}$$

Using the numerical values in Eqs. 2.47 and 2.48 we can write:

$$\begin{aligned} N_c &= 2.5 \times 10^{19} \left(\frac{m_c}{m_0}\right)^{3/2} \left(\frac{T}{300}\right)^{3/2} \text{ cm}^{-3} \\ N_v &= 2.5 \times 10^{19} \left(\frac{m_v}{m_0}\right)^{3/2} \left(\frac{T}{300}\right)^{3/2} \text{ cm}^{-3} \end{aligned} \tag{2.50}$$

The expressions of N_c and N_v differ only because there are differences in the values of the effective masses m_c and m_v. N_c and N_v are of the order of 10^{19} cm^{-3}, as reported in Table 2.2 in the case of Ge, Si, GaAs and InP. Equations Eqs. 2.47 and

	Ge	Si	GaAs	InP
N_c (cm^{-3})	1.04×10^{19}	3.2×10^{19}	4.4×10^{17}	5.7×10^{17}
N_v (cm^{-3})	5.0×10^{18}	1.83×10^{19}	8.2×10^{18}	1.2×10^{19}

Table 2.2: Values of N_c and N_v for Ge, Si, GaAs and InP at $T = 300$ K.

2.48 allow one to calculate the position of the Fermi level once the electron or hole densities are known. Indeed:

$$\begin{aligned} \mathcal{E}_F &= \mathcal{E}_c - k_B T \ln \frac{N_c}{n} \\ \mathcal{E}_F &= \mathcal{E}_v + k_B T \ln \frac{N_v}{p} \end{aligned} \tag{2.51}$$

We note that the product np is independent of the Fermi level and it is given by the following expression:

$$np = N_c N_v \exp\left(-\frac{\mathcal{E}_g}{k_B T}\right) \tag{2.52}$$

where $\mathcal{E}_g = \mathcal{E}_c - \mathcal{E}_v$ is the width of the bandgap. This is an important formula since it gives the relationship between electron and hole densities in a semiconductor at thermal equilibrium. Expression 2.52 is called *mass-action law*.

Intrinsic and extrinsic semiconductors

When the electrons in the conduction band are the result of thermal excitation from the valence band, the semiconductor is called *intrinsic*. An intrinsic semiconductor is also called undoped semiconductor, since it is a pure semiconductor without any dopant species. In an intrinsic semiconductor the number of electrons in the conduction band is equal to the number of holes in the valence band, since the holes

result from electrons which have been thermally excited to the conduction band. Therefore, we have:

$$n = p = n_i = \sqrt{N_c N_v}\, \exp\left(-\frac{\mathcal{E}_g}{2k_B T}\right) \tag{2.53}$$

where n_i is called intrinsic carrier density. The intrinsic carrier density can be written as follows

$$n_i = N_c \exp\left(-\frac{\mathcal{E}_c - \mathcal{E}_i}{k_B T}\right) = N_v \exp\left(-\frac{\mathcal{E}_i - \mathcal{E}_v}{k_B T}\right) \tag{2.54}$$

In an intrinsic semiconductor since $n = p = n_i$ and $N_c \approx N_v$ we have $(\mathcal{E}_c - \mathcal{E}_i) \approx (\mathcal{E}_i - \mathcal{E}_v)$ where $\mathcal{E}_i$ is the so-called intrinsic Fermi energy. Therefore, the Fermi level $\mathcal{E}_i$ is close to the middle of the energy gap $\mathcal{E}_i \approx (\mathcal{E}_c + \mathcal{E}_v)/2$. Also in the case of extrinsic semiconductors the intrinsic Fermi level is frequently used as a reference level. Indeed the expressions for the carrier densities n and p in an extrinsic semiconductor can be written in terms of the intrinsic carrier density and the intrinsic Fermi level:

$$n = n_i \exp\left(\frac{\mathcal{E}_F - \mathcal{E}_i}{k_B T}\right) \tag{2.55}$$

$$p = n_i \exp\left(\frac{\mathcal{E}_i - \mathcal{E}_F}{k_B T}\right) \tag{2.56}$$

Therefore the energy separation between the Fermi level and the intrinsic Fermi level is a measure of the departure of the semiconductor from an intrinsic material.

2.4 Mass-Action Law

The mass-action law can be deduced in a more general way by using thermodynamic considerations. We will assume, as we have done so far, that the semiconductor is in thermal equilibrium. This is a dynamical condition where every process is balanced by its inverse process. At thermal equilibrium some electrons are continuously excited by thermal energy from the valence band to the conduction band, while other electrons lose energy and from the conduction band return to the valence band. The excitation of an electron from the valence to the conduction band leads to the generation of a hole and an electron, while when an electron returns to the valence band an electron-hole recombination event is produced. It is reasonable to assume that the generation rate, G, depends on temperature T while, to first order, it does not depend on the number of carriers already present in the conduction and valence band. This last assumption is related to the extremely high number of bonds which can be broken by thermal energy, thus generating electrons in the conduction band and holes in the valence band. Therefore we can write:

$$G = f_1(T) \tag{2.57}$$

where $f_1(T)$ is a function which depends on the semiconductor characteristics and on temperature. On the other hand, the recombination rate, R, depends not only on temperature, but also on the density of electrons in the conduction band and of holes in the valence band, since each recombination process requires that an electron is present in the conduction band and a hole is present in the valence band. We can write:

$$R = npf_2(T) \tag{2.58}$$

where $f_2(T)$, as $f_1(T)$, is a function which depends on the semiconductor characteristics and on temperature. At thermal equilibrium the generation rate is equal to the recombination rate so that:

$$npf_2(T) = f_1(T) \tag{2.59}$$

or

$$np = \frac{f_1(T)}{f_2(T)} = f_3(T) \tag{2.60}$$

This is an important result: in a given semiconductor at thermal equilibrium the product of the hole and electron densities is a function only of temperature.
In an intrinsic semiconductor we have:

$$np = n_i^2 = f_3(T) \tag{2.61}$$

The intrinsic carrier density is a function of temperature since the thermal energy is the origin of the excitation of the electrons from the valence band to the conduction band. By using Eq. 2.53 we can write:

$$n_i^2 = N_c N_v \exp\left(-\frac{\mathcal{E}_g}{k_B T}\right). \tag{2.62}$$

The intrinsic carrier density is obviously a function of the energy gap since upon increasing the gap, fewer electrons can be excited into the conduction band by thermal energy. For silicon ($\mathcal{E}_g = 1.1$ eV) around room temperature n_i doubles for every 8-K increase in temperature. Equation 2.60 can be rewritten as:

$$np = n_i^2 \tag{2.63}$$

which is the mass-action law. This law holds for both intrinsic and extrinsic (i.e., doped, see next section) semiconductors.

2.5 Doped semiconductors

An intrinsic semiconductor is of limited use in applications. The situation is completely different when the material is doped, that is when particular chemical impurities are incorporated substitutionally in the semiconductor. This means that the impurities occupy lattice sites in place of the atoms of the pure semiconductor. For example, if an atom of Si, with four valence electrons, is substituted by an atom from group V, with five valence electrons, such as phosphorus, four of the valence electrons of the impurity atom fill bonds between the impurity atom and four adjacent Si atoms. The remaining fifth electron is not covalently bound to adjacent atoms and it is only weakly bound to the impurity atom by the excess positive charge of the nucleus. Therefore only a small amount of energy is required to break this weak bond so that the fifth electron is free to move in the material, thus contributing to electrical conduction. Since in this process the impurity atoms donate electrons to the semiconductor, they are called *donors* and the resulting semiconductor is called an $n-$type semiconductor.
The energy required to break the bond to a donor atom can be estimated by using a simple hydrogen-like model. Indeed, the fifth electron experiences the attractive

potential, $U(r)$, of the single net positive charge of the nucleus of the impurity atom, weakened by the polarization effects of the semiconductor material:

$$U(r) = -\frac{e^2}{4\pi\epsilon_o\epsilon_r r} \tag{2.64}$$

where ϵ_r is the relative dielectric constant of the semiconductor. Therefore we have essentially the same situation of the electron in the hydrogen atom, with two important differences: the electron mass is m^* (not the free electron mass) and $\epsilon_r \neq 1$, that is the Coulomb potential experienced by the electron is reduced by the factor ϵ_r. The Bohr theory of hydrogen can be easily modified to take into account these two differences. The ionization energy of atomic hydrogen is:

$$\mathcal{E}_H = \frac{e^4 m_0}{2(4\pi\epsilon_0\hbar)^2} = 13.6 \text{ eV} \tag{2.65}$$

The ionization energy of a donor impurity in the semiconductor, i.e., the binding energy, $\mathcal{E}_b$, of the electron to the core, can be obtained from Eq. 2.65 by using the dielectric constant of the semiconductor, $\epsilon = \epsilon_0\epsilon_r$ and by replacing m_0 with the effective mass m^*:

$$\mathcal{E}_b = \frac{e^4 m^*}{2(4\pi\epsilon_r\epsilon_0\hbar)^2} = 13.6\left(\frac{1}{\epsilon_r^2}\frac{m^*}{m_0}\right) \text{ eV} \tag{2.66}$$

The situation is schematically illustrated in Fig. 2.4. The effective mass which must

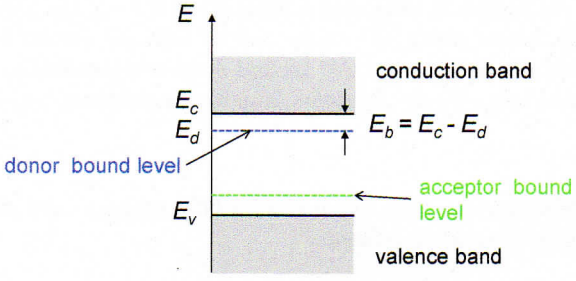

Figure 2.4: Schematic representation of the energy level of a donor impurity in a semiconductor. The binding energy is given by $\mathcal{E}_b = \mathcal{E}_c - \mathcal{E}_d$. Figure also shows an acceptor bound level.

be used in this case is the *conductivity effective mass*, m_σ^*, whose value depends on how electrons respond to an external potential. For direct bandgap semiconductors m_σ^* is simply the effective mass. For semiconductors such as Si, the conductivity mass is given by the following equation:

$$m_\sigma^* = 3\left(\frac{2}{m_t} + \frac{1}{m_l}\right)^{-1}. \tag{2.67}$$

Exercise 2.2 Calculate the binding energy of the donor states in GaAs and Si. Numerical data for Si: conduction band effective mass $m_l = 0.98\ m_0$, $m_t = 0.2\ m_0$, $\epsilon_r = 13.2$. Numerical data for GaAs: conduction band effective mass $m_c = 0.067\ m_0$, $\epsilon_r = 11.9$.

In the case of GaAs we have:

$$\mathcal{E}_b = 13.6 \left(\frac{1}{\epsilon_r^2} \frac{m^*}{m_0} \right) = 5.2 \text{ meV}$$

In the case of Si we have first to calculate the conductivity effective mass:

$$m_\sigma^* = 3 \left(\frac{2}{m_t} + \frac{1}{m_l} \right)^{-1} = 0.26\, m_0.$$

The ionization energy of the donor is given by $\mathcal{E}_b = 25$ meV, which is only about 3% of the energy gap. ■

We can also obtain a simple estimation of the mean distance of the electron from the donor atom. The Bohr radius of the ground state of hydrogen is related to the ionization energy by the following formula:

$$\mathcal{E}_H = \frac{e^2}{2(4\pi\epsilon_0 R_H)} \tag{2.68}$$

so that

$$R_H = \frac{4\pi\epsilon_0\hbar^2}{m_0 e^2} = 0.53 \text{ Å} \tag{2.69}$$

The Bohr radius of the donor can be obtained from the previous expression by considering the relative dielectric constant of the semiconductor and the effective mass:

$$R_d = \frac{4\pi\epsilon_r\epsilon_0\hbar^2}{m^* e^2} = 0.53 \left(\epsilon_r \frac{m_0}{m^*} \right) \text{ Å} \tag{2.70}$$

The simple hydrogen-like model we have used so far predicts that the binding energy

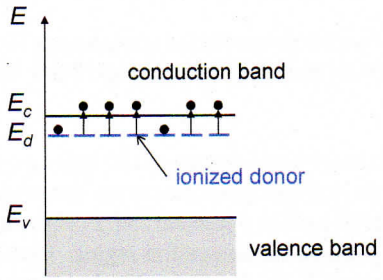

Figure 2.5: Donor states are neutral when occupied by an electron and positively charged when empty.

of a donor state depends only on the semiconductor crystal (through ϵ_r and m^*) and not on the particular dopant atom. In a real situation the ionization energy depends also on the donor atom as a result of a small distortion of the atomic potential. For example, measured values of the donor ionization energies in Si give $\mathcal{E}_b = 45$ meV for P, $\mathcal{E}_b = 49$ meV for As, $\mathcal{E}_b = 39$ meV for Sb. In Ge $\mathcal{E}_b = 12$ meV for P, $\mathcal{E}_b = 12.7$ meV for As, $\mathcal{E}_b = 9.6$ meV for Sb. The Bohr radii of the donor states obtained from Eq. 2.70 are (132)(0.53) Å $\simeq$ 70 Å for Ge, (45)(0.53) Å $\simeq$ 24 Å for Si and (193)(0.53) Å $\simeq$ 102 Å for GaAs, which are quite large values, which lead to overlap of the

dopant orbitals even at relatively low donor concentrations compared to the number of atoms of the host semiconductor. The orbital overlap leads to the generation of an impurity band formed from the donor states.
Due to the small ionization energy of the donor states, thermal energy is large enough to excite electrons from the bound donor states to semiconductor conduction band already at temperatures >150 K. When an electron is excited in the conduction band, a fixed and positively charged atom core is left in the semiconductor in the dopant site. Therefore the allowed energy states introduced by donors are neutral when occupied by an electron and positively charged when they are empty, as schematically shown in Fig. 2.5.
When a semiconductor is doped with donor atoms the number of electrons in the conduction band is much larger than the number of holes in the valence band. Therefore, in a $n-$type semiconductor the electrons are majority carriers and the holes are the minority carriers. Figure 2.6 shows the electron density in the conduction

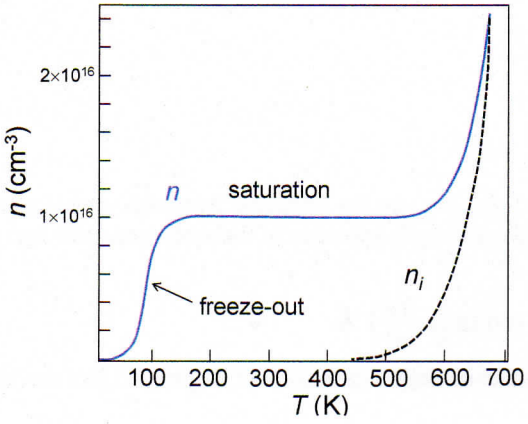

Figure 2.6: Electron concentration versus temperature for silicon doped with 1×10^{16} arsenic atoms cm^{-3} (solid blue line). The dashed black line shows the intrinsic carrier density, n_i.

band of Si as a function of temperature in the case of Si doped with $N_d = 1 \times 10^{16}$ arsenic atoms cm^{-3}. At very low temperature the donor electrons are tied to the donor sites: this effect is called *carrier freeze-out*. At $T > 150$ K all donors are ionized and $n \simeq N_d$. This region is called the saturation region. Upon increasing the temperature up to a few hundreds of K ($\sim$ 600 K in Si) the thermal energy is able to ionize many electrons from silicon-silicon bonds so that the intrinsic carrier population starts to increase and becomes the dominant source of electrons in the conduction band. Semiconductor devices usually work in the saturation region, where the free carrier density is almost independent of temperature and it is approximately equal to the density of the dopant atoms. Semiconductor devices cannot operate in the high-temperature intrinsic region since it is not possible to control the intrinsic carrier density by applying external bias. For this reason, the devices which must work at high temperature are based on the use of large bandgap semiconductors in order to increase the temperature required to ionize the bonds between the semiconductor atoms.

If a semiconductor with four valence electrons as Si is doped with impurity atoms

with three valence electrons (such as B, Al, Ga and In), the three electrons fill three of the four covalent bonds of Si and one bond remains vacant. This bond can be filled by an electron coming from a nearby covalent bond, so that the vacant bond is moved carrying with it a positive charge. In this case the trivalent dopant atoms are called *acceptors* since they accept electrons from the valence band of the host semiconductor to complete the four covalent bonds with neighbour atoms, thus leaving holes in the valence band. The material is called $p-$type semiconductor, since most of the conduction is carried by positive charges (holes). The energy required to excite an electron from the valence band to the vacant bond is typically of the same order of magnitude of the binding energy of a donor state. For example the acceptor energies $\mathcal{E}_a$ of trivalent atoms in Si are $\mathcal{E}_a = 45$ meV for B, $\mathcal{E}_a = 57$ meV for Al, $\mathcal{E}_a = 65$ meV for Ga. The corresponding acceptor state is represented by an energy level slightly above the top of the valence band, as shown in Fig. 2.4. An acceptor level is neutral when empty and negatively charged when it is occupied by an electron. The hydrogen-like model used in the case of donor impurities can be applied also in the case of acceptor atoms, but its application is complicated by the degeneracy at the valence bandedge and it does not give accurate results.

In the case of compound semiconductors like GaAs, tetravalent impurities may substitutionally replace either Ga or As. For example, Si acts as a donor impurity when it replaces Ga, whereas it acts as an acceptor impurity when it substitutes for As. A different situation is encountered for example when tin is used in GaAs since it replaces almost exclusively Ga in GaAs thus leading to an effective $n-$type doping. $n-$type GaAs can be obtained by using impurities from group VI of the periodic table, such as tellurium, selenium or sulphur. $p-$type doping of GaAs can be obtained by using impurities from group II, like zinc or cadmium.

2.5.1 Carrier density and Fermi energy in doped semiconductors

We assume to have a semiconductor doped with N_d donor atoms per unit volume and N_a acceptors per unit volume. Moreover we assume that all dopant are ionized and that $N_d > N_a$. Due to the charge neutrality of the material we have:

$$n + N_a = p + N_d \tag{2.71}$$

Using the mass-action law ($np = n_i^2$) in the previous equation we obtain

$$n - \frac{n_i^2}{n} = N_d - N_a \tag{2.72}$$

which may be solved for the electron concentration n:

$$n = \frac{N_d - N_a}{2} + \left[\left(\frac{N_d - N_a}{2}\right)^2 + n_i^2\right]^{1/2} \tag{2.73}$$

If we consider a n-type semiconductor $N_d - N_a > 0$ and $N_d - N_a \gg n_i$. Therefore from Eq. 2.73 we have:

$$n \approx N_d - N_a \tag{2.74}$$

$$p = \frac{n_i^2}{n} \approx \frac{n_i^2}{N_d - N_a} \tag{2.75}$$

From the last equation we see that the electron density depends on the net excess of ionized donors over acceptors. Thus a p-type material containing N_a acceptors can

be converted into an n-type material by adding an excess of donors so that $N_d > N_a$. For silicon at room temperature, $n_i = 1.45 \times 10^{10}$ cm^{-3} while the net donor density in n-type silicon is typically about 10^{15} cm^{-3} or greater, so that $(N_d - N_a) \gg n_i$. Thus, for $N_d - N_a = 10^{15}$ cm^{-3} we have $n \approx N_d - N_a$ and $p = 2 \times 10^5$ cm^{-3}, so that the minority-carrier concentration is nearly 10 orders of magnitude below the majority-carrier population. In general, the concentration of one type of carrier is many orders of magnitude greater than that of the other in extrinsic semiconductors.

By using Eq. 2.51 it is possible to calculate the position of the Fermi level:

$$\mathcal{E}_F = \mathcal{E}_c - k_B T \log \frac{N_c}{N_d - N_a} \tag{2.76}$$

In the same way, assuming a $p-$type doping ($N_a > N_d$) we have:

$$\mathcal{E}_F = \mathcal{E}_v + k_B T \log \frac{N_v}{N_a - N_d}. \tag{2.77}$$

2.5.2 Degenerate semiconductors

The previous equations show that, if the doping level is less than the effective density of states (N_c and N_v) the semiconductor remains non-degenerate, i.e., the Fermi level is located in the energy gap. When the semiconductor is heavily doped, i.e., when N_d or N_a approaches the effective density of states in the conduction or valence band (therefore $\sim 10^{18} - 10^{19}$ cm^{-3} for typical semiconductors), we cannot employ the approximation used in the case of non-degenerate semiconductors. Since we cannot approximate the Fermi-Dirac distribution function by the Maxwell-Boltzmann distribution function because it is not possible to neglect the effects of Pauli exclusion principle, Eqs.2.47 and 2.48 are no longer valid. A semiconductor with very high doping levels ($N_d \gtrsim N_c$ or $N_a \gtrsim N_v$) is called *degenerate semiconductor* because the corresponding Fermi level is located within the conduction or the valence band. In the case of heavy doping other effects must be taken into account, which are completely negligible for non-degenerate semiconductors. In the case of small dopant concentration the mean distance between dopant atoms is typically much larger than the extension of their wavefunction, and the overlapping between the wavefunctions of the dopant atoms is very small. For this reason the energy levels of the dopants can be considered as discrete levels, as shown in Fig. 2.7(a). In heavily-doped semiconductors the separation between dopant atoms reduces to a few angstroms, so that impurity bands are generated, as shown in Fig. 2.7(b).

Another important effect is related to electron-electron interaction, when a large number of electrons are introduced in the conduction band, or when a large number of holes is introduced in the valence band. Electron-electron interaction produces a narrowing of the semiconductor energy gap. For example, in the case of bulk GaAs the energy gap narrowing is given by (from H.C. Casey Jr. and F. Stern, J. Appl. Phys. **47**, 631 (1976)):

$$\Delta\mathcal{E}_g = \mathcal{E}_g - \mathcal{E}_{g0} = -1.6 \times 10^{-8}(p^{1/3} + n^{1/3}) \text{ eV}, \tag{2.78}$$

where $\mathcal{E}_{g0}$ is the energy gap at zero doping, n and p are the electron and hole densities measured in cm^{-3}. Assuming a donor density of $\sim 10^{18}$ cm^{-3} an energy gap reduction of $\Delta\mathcal{E}_g \simeq 16$ meV is obtained. In the case of silicon the band gap narrowing

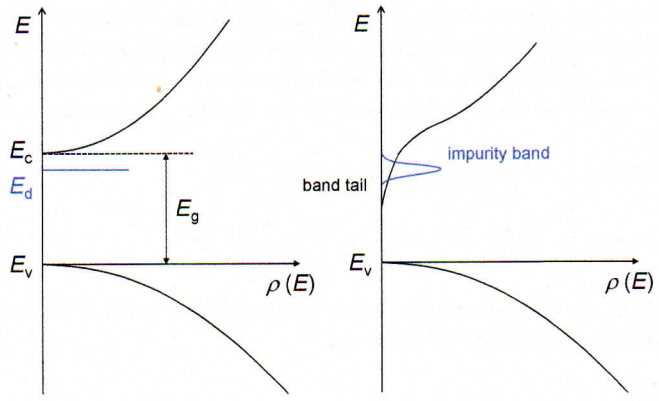

Figure 2.7: Schematic energy band diagram for (a) a non-degenerate semiconductor and for (b) a heavily doped semiconductor. $\rho(\mathcal{E})$ is the density of states.

can be well reproduced by the following expression, where T is the temperature in K:

$$\Delta\mathcal{E}_g \simeq -22.5\left(\frac{N_d}{10^{18}}\frac{300}{T}\right)^{1/2} \text{ meV}, \tag{2.79}$$

Another effect it is worth to mention is related to the fact that a very high dopant concentration disrupts the perfect periodicity of the crystal lattice, thus leading to the formation of a band tail, as shown in Fig. 2.7(b).

2.6 Quasi-Fermi levels in non-equilibrium systems

So far we have always assumed semiconductors in thermal equilibrium. There are situations where the material is in a non-equilibrium condition. Indeed, all semiconductor devices work under non-equilibrium conditions caused, for example, by the application of an external voltage or by exposing the semiconductor to light with photon energy greater than the energy gap. In all these cases electron and hole populations are no longer in thermodynamic equilibrium with each other. Let us assume that electrons have been excited into the conduction band by an external excitation. Typically, the intraband relaxation processes, with a characteristic relaxation time of the order of ~ 1 ps, due to electron-phonon scattering processes, is much faster than the interband relaxation processes, with a characteristic relaxation time of ~ 1 ns, related to electron-hole recombination processes. Due to the very different temporal scales of the intra- and inter-band relaxation processes a thermal equilibrium condition can be reached in a few picoseconds both in the conduction and in the valence band. The semiconductor is still in a non-equilibrium situation as a whole: conduction and valence band are independently in thermal equilibrium. For this reason we can introduce two Fermi-Dirac distribution functions, $f_c(\mathcal{E})$ and $f_v(\mathcal{E})$, for the conduction and valence band, respectively, given by:

$$f_c(\mathcal{E}) = \frac{1}{1+\exp(\mathcal{E}-\mathcal{E}_{Fc})/k_BT} \tag{2.80}$$

$$f_v(\mathcal{E}) = \frac{1}{1+\exp(\mathcal{E}-\mathcal{E}_{Fv})/k_BT} \tag{2.81}$$

where $\mathcal{E}_{Fc}$ and $\mathcal{E}_{Fv}$ are the energies of the so-called *quasi-Fermi levels* for the conduction and valence band, respectively. If Maxwell-Boltzmann statistics can be applied ($n \ll N_c$, $p \ll N_v$) the electron and hole densities can be calculated by using Eqs. 2.47-2.48 and 2.55-2.56, replacing $\mathcal{E}_F$ with $\mathcal{E}_{Fc}$ and $\mathcal{E}_{Fv}$, respectively, so that:

$$n = N_c \exp\left(-\frac{\mathcal{E}_c - \mathcal{E}_{Fc}}{k_B T}\right) = n_i \exp\left(\frac{\mathcal{E}_{Fc} - \mathcal{E}_i}{k_B T}\right), \tag{2.82}$$

$$p = N_v \exp\left(-\frac{\mathcal{E}_{Fv} - \mathcal{E}_v}{k_B T}\right) = n_i \exp\left(\frac{\mathcal{E}_i - \mathcal{E}_{Fv}}{k_B T}\right), \tag{2.83}$$

The position of the quasi-Fermi levels can be obtained in terms of the intrinsic Fermi level, $\mathcal{E}_i$, and the intrinsic carrier density:

$$\mathcal{E}_{Fc} = \mathcal{E}_i + k_B T \log\left(\frac{n}{n_i}\right) \tag{2.84}$$

$$\mathcal{E}_{Fv} = \mathcal{E}_i - k_B T \log\left(\frac{p}{n_i}\right) \tag{2.85}$$

It is not surprising that under non-equilibrium conditions the product np is not equal to the thermal equilibrium value n_i^2. Indeed, by using Eqs. 2.84 and 2.85 we obtain:

$$np = n_i^2 \exp\left(\frac{\mathcal{E}_{Fc} - \mathcal{E}_{Fv}}{k_B T}\right). \tag{2.86}$$

Therefore, the energy separation between the two quasi-Fermi levels can be considered as a measure of the departure of the semiconductor from the thermal equilibrium situation. At thermal equilibrium ($\mathcal{E}_{Fc} = \mathcal{E}_{Fv}$) the previous equation reduces to the usual expression: $np = n_i^2$.

Exercise 2.3 Calculate the energy of the quasi-Fermi levels in GaAs at $T = 0$ K and at $T = 300$ K for an injected electron density $n = 2.86 \times 10^{18}$ cm^{-3} and for an injected hole density $p = 2.86 \times 10^{18}$ cm^{-3}. Numerical values: $m_c = 0.067\, m_0$, $m_{hh} = 0.45\, m_0$, $m_{lh} = 0.082\, m_0$.

At $T = 0$ K the calculation of the quasi-Fermi energy is particularly simple since, in this case, all allowed states with $\mathcal{E} < \mathcal{E}_{Fc}$ are filled by electrons, while all allowed states with $\mathcal{E} > \mathcal{E}_{Fc}$ are empty, so that the carrier density can be calculated analytically:

$$n = \frac{1}{2\pi^2}\left(\frac{2m_c}{\hbar^2}\right)^{3/2} \int_{\mathcal{E}_c}^{\mathcal{E}_{Fc}} \sqrt{\mathcal{E} - \mathcal{E}_c}\, d\mathcal{E} = \frac{1}{3\pi^2}\left(\frac{2m_c}{\hbar^2}\right)^{3/2} (\mathcal{E}_{Fc} - \mathcal{E}_c)^{3/2}$$

so that:

$$\mathcal{E}_{Fc} - \mathcal{E}_c = \frac{\hbar^2}{2m_c}(3\pi^2 n)^{2/3} = 109 \text{ meV}.$$

We can already anticipate that, since in this case $\mathcal{E}_{Fc} - \mathcal{E}_c \gg k_B T$ at room temperature, the previous result can be considered a good approximation of the position of the quasi-Fermi level also at room temperature.

In the same way for the valence band one obtains:

$$\mathcal{E}_v - \mathcal{E}_{Fv} = \frac{\hbar^2}{2m_v}(3\pi^2 p)^{2/3} = 16 \text{ meV},$$

where the effective mass m_v is given by Eq. 2.29: $m_v = (m_{hh}^{3/2} + m_{lh}^{3/2})^{2/3} = 0.47\, m_0$. Since in this case $\mathcal{E}_v - \mathcal{E}_{Fv} < k_B T$ (at room temperature), we expect that the result of Eq. 2.87 is valid only at very low temperatures.
At $T = 300$ K the electron density in the conduction band can be calculated by using Eq. 2.42. We have first to calculate the effective density of states in the conduction band:

$$N_c = \frac{1}{4}\left(\frac{2m_c k_B T}{\pi\hbar^2}\right)^{3/2} = 4.4 \times 10^{17} \text{ cm}^{-3}.$$

From Eq. 2.42 one obtains:

$$F_{1/2}(y) = \frac{n}{N_c} = 6.5$$

From Fig. 2.3 it is possible to obtain the corresponding y value: $y \simeq 4$, so that:

$$\mathcal{E}_{Fc} - \mathcal{E}_c \simeq 4k_B T = 103.5 \text{ meV}$$

Thus confirming that the result obtained at $T = 0$ K was a good approximation of the correct value of the quasi-Fermi energy in the conduction band. Since $y \gg 1$, we can use the approximation of the Fermi function given by Eq. 2.44:

$$y = \left[\frac{3}{4}\sqrt{\pi} F_{1/2}(y)\right]^{2/3} \simeq 4.21,$$

thus giving $\mathcal{E}_{Fc} - \mathcal{E}_c \simeq 4.21 k_B T = 109$ meV, which represents a good approximation of the correct result.
The same procedure can be adopted in the case of hole injection. In this case:

$$N_v = \frac{1}{4}\left(\frac{2m_v k_B T}{\pi\hbar^2}\right)^{3/2} = 8.16 \times 10^{18} \text{ cm}^{-3},$$

$$F_{1/2}(y) = \frac{p}{N_v} = 0.35$$

where $y = \mathcal{E}_v - \mathcal{E}_{Fv}/k_B T$. From Fig. 2.3 it is possible to obtain the corresponding y value: $y \simeq -0.95$, so that:

$$\mathcal{E}_v - \mathcal{E}_{Fv} = y k_B T \simeq -24.6 \text{ meV},$$

thus showing that, as already anticipated, the result obtained at $T = 0$ does not represent a reasonable approximation of the correct one. Since y is negative we can use the exponential approximation of the Fermi function: $F_{1/2}(y) \approx e^y$, so that $y = \log[F_{1/2}(y)] = -1.05$ and $\mathcal{E}_v - \mathcal{E}_{Fv} = -27$ meV. ■

2.7 Charge transport in semiconductors

In this section we will briefly discuss the charge transport processes in semiconductors. When the material is in thermal equilibrium the electrons in the conduction band move in random motion as almost free particles: the influence of the crystal lattice is contained in the effective mass. According to the Boltzmann's law of equi-distribution of kinetic energy per degrees of freedom of particles, the average kinetic energy of a particle is $(1/2)k_BT$ per each degree of freedom. As a first order approximation, the average thermal kinetic energy of the electrons can be written as:

$$\frac{1}{2}m^*v_{th}^2 = \frac{3}{2}k_BT, \tag{2.87}$$

where m^* is the effective mass of electrons in the conduction band. For example, in the case of silicon at $T = 300$ K the thermal velocity v_{th} turns out to be of the order of $\sim 10^7$ cm/s. In a perfect lattice no scattering events are present and the effects related to the lattice are taken into account in the effective mass. In a real lattice various scattering mechanisms are present. Indeed, at $T > 0$ K the lattice atoms vibrate about their equilibrium position. In quantum mechanics this process is described in terms of electron-phonon scattering. In a real crystal we can also have scattering at ionized impurities, which is relevant at high dopant densities; electron-electron scattering, important at high carrier density; scattering at crystal defects. At thermal equilibrium the motion of electrons is completely random so that the net current in any direction is zero. The collisions with lattice determine a transfer of energy between the electrons and the atoms, which form the lattice. The average time interval between consecutive collisions, τ_n, is called mean scattering time and it is of the order of 1 ps. Therefore, the average distance between consecutive collisions, called mean free path, is roughly given by $\ell_n = v_{th}\tau_n \approx 0.1$ μm.
If an electric field is applied to the crystal the electrons are accelerated in the opposite direction and an overall motion of carriers in the direction of the electric field is produced, superposed to the random thermal motion. If the applied field is not too high, we can assume that the mean scattering time is not notably changed by the application of the field. The average net velocity in the direction of the field is called drift velocity, v_d. In a steady-state condition, the momentum acquired by the electron between consecutive collisions is lost to the lattice in the collision. The momentum gained by an electron between collisions is equal to the impulse of the force acting on the electron:

$$m_n^* v_d = (-eE)\tau_n \tag{2.88}$$

so that the drift velocity can be written as:

$$v_d = -\frac{e\tau_n}{m_n^*}E = -\mu_n E \tag{2.89}$$

where

$$\mu_n = \frac{e\tau_n}{m_n^*} \tag{2.90}$$

is the electron mobility and describes how an electron responds to an applied electric field. The drift of electrons induced by the applied field gives rise to a current flow, with a current density given by:

$$J_n = -nev_d = ne\mu_n E. \tag{2.91}$$

We can repeat the same discussion for holes in the valence band. The hole mobility can be written as:

$$\mu_p = \frac{e\tau_p}{m_p^*} \tag{2.92}$$

and the total current density can be written as:

$$J = J_n + J_p = (ne\mu_n + pe\mu_p)E = \sigma E \tag{2.93}$$

where

$$\sigma = ne\mu_n + pe\mu_p \tag{2.94}$$

is the conductivity of the semiconductor. In an $n-$type semiconductor $J \simeq J_n$, while in a $p-$type semiconductor $J \simeq J_p$.
Various scattering processes are present in a semiconductor, each characterized by a given mean scattering time τ_i and each giving rise to a different mobility μ_i. If we assume that the probability that a carrier is scattered in a time interval dt by the scattering process i is dt/τ_i, the total scattering probability in the same time interval can be written as:

$$\frac{dt}{\tau_c} = \sum_i \frac{dt}{\tau_i}. \tag{2.95}$$

Therefore, the mobility $\mu = e\tau_c/m^*$ can be written as (*Matthiessen's rule*):

$$\frac{1}{\mu} = \sum_i \frac{1}{\mu_i}. \tag{2.96}$$

It is therefore evident that the value of the mobility is determined mainly by the scattering process characterized by the smallest scattering time. The mobility depends on total dopant concentration as shown in Fig. 2.8, which displays electron and hole mobility for Si at 300 K as a function of the total dopant density. In the case of

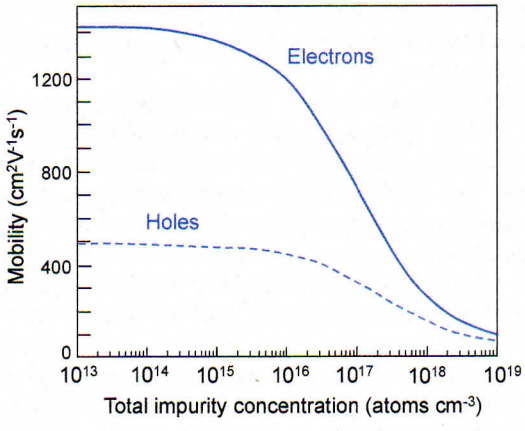

Figure 2.8: Electron and hole mobility versus doping density for silicon at room temperature.

low dopant density (of the order of 10^{15} cm^{-3}) the mobility resulting from lattice phonon scattering is smaller than that due to ionized impurity scattering, therefore for dopant densities smaller than $\sim 10^{15}$ cm^{-3} (in Si) the electron and hole mobility is almost constant, as clearly shown in Fig. 2.8. In other words, at low doping lattice (phonon) scattering dominates. Upon increasing the dopant density, the mobility related to impurity scattering becomes comparable to that related to lattice scattering and the total mobility decreases. It is evident that in a compensated semiconductor (i.e., when both donor and acceptor atoms are introduced) while the carrier density depends on the difference between the dopant concentrations, $N_d - N_a$, the scattering depends on the total dopant density, $N_d + N_a$. Therefore compensated semiconductors have a lower carrier mobility for the same net dopant density than that of an un-compensated semiconductor (i.e., with only a single dopant type). A very good approximation of measured dependence of μ vs dopant density in Si is given by the following expression:

$$\mu = \mu_{min} + \frac{\mu_{max} - \mu_{min}}{1 + (N/N_{ref})^{\alpha}} \tag{2.97}$$

where N is the total dopant density and the parameters μ_{min}, μ_{max}, N_{ref} and α depend on the dopant atom.

The different scattering processes show different temperature dependences, thus leading to a complex dependence of the total mobility vs temperature. Without entering into details, upon increasing the temperature, lattice vibrations increase, thus increasing the lattice scattering component and reducing the corresponding mobility:

$$\mu_{lattice} \propto T^{-n}, \qquad n \sim 1.5 \tag{2.98}$$

In contrast, the impurity scattering component is reduced at high temperature since the carriers move faster thus remaining near the impurities for shorter time:

$$\mu_{impurity} \propto T^{n}, \qquad n \sim 1.5 \tag{2.99}$$

Therefore at higher temperatures lattice scattering tends to dominate over impurity scattering and the mobility decreases. At lower temperature the reverse situation is found and the mobility increases with temperature. This behaviour gives rise to a maximum of the curve $\mu(T)$.

Before closing this section we note that the linear dependence of the drift velocity vs applied electric field, $v_d = \mu E$, is valid only for small values of the applied field, a shown in Fig. 2.9, which displays the drift velocity of electrons and holes as a function of the electric field amplitude in silicon. In the derivation of the linear relation between v_d and E we have assumed that the average collision time τ does not depend on the applied electric field. Note that for low electric field amplitudes the drift velocity is much smaller than the thermal velocity. In the case of Si $\mu_n \approx 1400$ cm^{2}V^{-1}s^{-1}, assuming a field amplitude $E = 100$ Vcm^{-1}, the drift velocity is $v_d \approx 1.4 \times 10^5$ cm/s, while $v_{th} \approx 10^7$ cm/s, so that the applied electric field does not significantly alter the total electron velocity. Upon increasing the applied electric field the electron kinetic energy increases and overcomes the thermal energy. In this regime the high-energy electrons are called *hot electrons*. The energy transferred from the hot electrons to the lattice is high. When the hot electron density reaches a critical value, a new scattering process comes into play: the scattering with

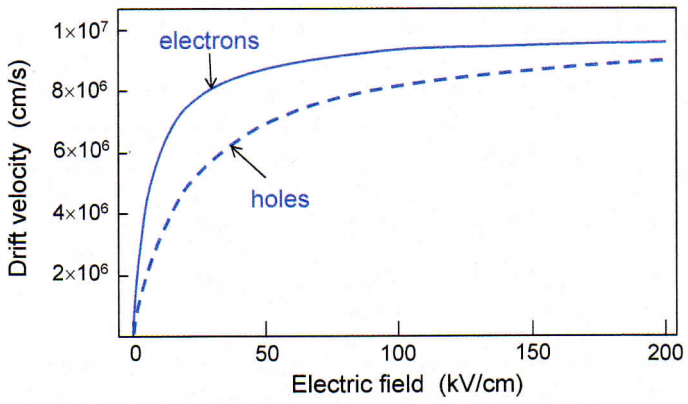

Figure 2.9: Drift velocity of electrons and holes as a function of the electric field amplitude in silicon.

high-energy or *optical phonons*. In this situation the drift velocity cannot increase linearly with E and reaches a saturation value at high E. A very good approximation of the dependence of the drift velocity on the applied electric field in Si is given by the following empirical expression:

$$v_d = v_\ell \frac{E}{E_c} \left[\frac{1}{1 + (E/E_c)^\beta} \right]^{1/\beta} \tag{2.100}$$

where v_ℓ, E_c and β are suitable parameters (e.g., for electrons in Si at 300 K $v_\ell = 1.07 \times 10^7$ cm/s, $E_c = 6.91 \times 10^3$ V/cm and $\beta = 1.11$).

2.8 Diffusion current

In a semiconductor there is a second component of current, which is generated by a nonuniform spatial distribution of carrier density: this current component, which is particularly important in semiconductors, is called diffusion current. Diffusion of carriers occurs from regions of high density towards regions of low carrier density. In other words, gradients of the carrier densities, ∇n or ∇p, generate electron or hole currents, according to the Fick's law:

$$\mathbf{J}_n = eD_n \nabla n \tag{2.101}$$

$$\mathbf{J}_p = -eD_p \nabla p \tag{2.102}$$

where D_n and D_p are the diffusion coefficients for electrons and holes, respectively. The current densities $\mathbf{J}_n$ and $\mathbf{J}_p$ in a semiconductor with an applied electric field are given by the sum of the drift current and of the diffusion current:

$$\mathbf{J}_n = e\mu_n n\mathbf{E} + eD_n \nabla n \tag{2.103}$$

$$\mathbf{J}_p = e\mu_p p\mathbf{E} - eD_p \nabla p \tag{2.104}$$

and the total current density is given by $\mathbf{J} = \mathbf{J}_n + \mathbf{J}_p$. D_n and D_p are given by the Einstein relations:

$$\frac{D_n}{\mu_n} = \frac{D_p}{\mu_p} = \frac{k_B T}{e} \tag{2.105}$$

Table 2.3 shows the mobility and diffusion coefficients at $T = 300$ K for Si, GaAs and Ge corresponding to the typical mobility values at the same temperature and low dopant density.

Equations 2.103 and 2.104 can be written in a different way in terms of the quasi-

	μ_n (cm^2/Vs)	D_n (cm^2/s)	μ_p (cm^2/Vs)	D_p (cm^2/s)
Si	$\leq$1400	$\leq$36	$\leq$450	$\leq$12
Ge	$\leq$3900	$\leq$100	$\leq$1900	$\leq$50
GaAs	$\leq$8500	$\leq$200	$\leq$400	$\leq$10
InP	$\leq$5400	$\leq$130	$\leq$200	$\leq$5
InAs	$\leq 4 \times 10^4$	$\leq$1000	$\leq$500	$\leq$13
InSb	$\leq 7.7 \times 10^4$	$\leq 2 \times 10^3$	$\leq$850	$\leq$22

Table 2.3: Typical mobility and diffusion coefficient at $T = 300$ K and low dopant concentration for a few semiconductors.

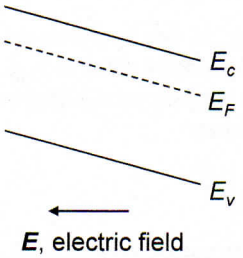

Figure 2.10: Schematic energy-band of an $n-$type semiconductor with an applied electric field $\mathbf{E}$.

Fermi levels. Quasi-Fermi levels have to be used since when current is flowing the semiconductor is not in thermal equilibrium. The potential, ϕ, is related to the applied electric field by the following relation:

$$\mathbf{E} = -\nabla\phi, \tag{2.106}$$

since the electron energy, $\mathcal{E}_c - \mathcal{E}_0$ (where $\mathcal{E}_0$ is an arbitrary fixed energy level) is given by:

$$\mathcal{E}_c - \mathcal{E}_0 = -e\phi, \tag{2.107}$$

the electric field can be written in terms of the gradient of the energy $\mathcal{E}_c$:

$$\mathbf{E} = \frac{1}{e}\nabla\mathcal{E}_c. \tag{2.108}$$

In the same way we can write:

$$\mathbf{E} = \frac{1}{e}\nabla\mathcal{E}_c = \frac{1}{e}\nabla\mathcal{E}_v = \frac{1}{e}\nabla\mathcal{E}_i \tag{2.109}$$

Therefore, when an electric field is present in the semiconductor, the conduction, valence and intrinsic energies vary with position, as schematically shown in Fig. 2.10.

If Maxwell-Boltzmann statistics can be used ($n \ll N_c$, $p \ll N_v$), the electron and hole densities are given by Eqs. 2.82 and 2.83. The electron current density is given by Eq. 2.103. Δn can be calculated by using Eq. 2.82:

$$\nabla n = \frac{n}{k_B T}(\nabla \mathcal{E}_{Fc} - \nabla \mathcal{E}_c) = \frac{n}{k_B T}\nabla \mathcal{E}_{Fc} - \frac{ne}{k_B T}\mathbf{E} \tag{2.110}$$

where we have used Eq. 2.108. Therefore, by using Eq. 2.110 and the Einstein relation, J_n can be written as:

$$\mathbf{J}_n = \mu_n n \nabla \mathcal{E}_{Fc} \tag{2.111}$$

In the same way for holes we obtain:

$$\mathbf{J}_p = \mu_p p \nabla \mathcal{E}_{Fv}. \tag{2.112}$$

Therefore, the total current density for electrons and holes is proportional to the gradient of the quasi-Fermi level of the respective carrier type.

3. Basic concepts of quantum mechanics

3.1 Quantum mechanics fundamentals

In quantum mechanics it is assumed that a particle can be described by a wavefunction $\psi(\mathbf{r}, t)$, which satisfies the Schrödinger equation:

$$H\psi(\mathbf{r}, t) = i\hbar \frac{\partial \psi(\mathbf{r}, t)}{\partial t} \tag{3.1}$$

where H is the Hamiltonian operator, which contains the information about the system of interest. It is assumed that $\psi(\mathbf{r}, t)$ is normalized:

$$\int \psi^* \psi \, d\tau = 1 \tag{3.2}$$

where the asterisk indicates that complex conjugate of ψ is taken and the integral is calculated over all space. We will generally use the Dirac notation, so that the integral $\int \phi^* \psi \, d\tau$, where $\phi(\mathbf{r}, t)$ and $\psi(\mathbf{r}, t)$ are two wavefunctions, can be represented as:

$$\int \phi^* \psi \, d\tau = \langle \phi | \psi \rangle \tag{3.3}$$

Each wavefunction, ψ, can be represented by a *ket* vector $|\psi\rangle$. There is a one-to-one correspondence between a ket vector and another vector, $\langle \psi |$, called *bra* vector, such that the product $\langle \psi | \psi \rangle$ represents the integral $\int \psi^* \psi \, d\tau$. Equation 3.3 defines the scalar product (or inner product) between the two wavevectors $|\phi\rangle$ and $|\psi\rangle$. The scalar product has the following properties:

$$\langle \phi | \psi \rangle = \langle \psi | \phi \rangle^*, \tag{3.4}$$

$$\langle \phi | \phi \rangle \text{ is real and positive (it is zero only if } |\phi\rangle = 0) \tag{3.5}$$

Indeed:

$$\langle\phi|\psi\rangle = \int \phi^*\psi \, d\tau = \left(\int \psi^*\phi \, d\tau\right)^* = \langle\psi|\phi\rangle^*, \tag{3.6}$$

$$\langle\phi|\phi\rangle = \int \phi^*\phi \, d\tau = \int |\phi|^2 \, d\tau \geq 0. \tag{3.7}$$

Moreover:

$$\langle\phi|\alpha\psi_1 + \beta\psi_2\rangle = \alpha\langle\phi|\psi_1\rangle + \beta\langle\phi|\psi_2\rangle, \tag{3.8}$$

$$\langle\alpha\phi_1 + \beta\phi_2|\psi\rangle = \alpha^*\langle\phi_1|\psi\rangle + \beta^*\langle\phi_2|\psi\rangle. \tag{3.9}$$

By using the Dirac notation the Schrödinger equation 3.1 can be written as:

$$H|\psi\rangle = i\hbar\frac{\partial|\psi\rangle}{\partial t}, \tag{3.10}$$

with the normalization condition:

$$\langle\psi|\psi\rangle = 1. \tag{3.11}$$

Let us first consider an isolated system, without any perturbation or interaction, described by the unperturbed Hamiltonian H_0, which is assumed time-independent. For this system there is a set of stationary states $|\phi_k\rangle$, which can be calculated by solving the Schrödinger equation:

$$H_0|\phi_k\rangle = i\hbar\frac{\partial|\phi_k\rangle}{\partial t}. \tag{3.12}$$

A solution of Eq. 3.12 can be written as:

$$|\phi_k(\mathbf{r},t)\rangle = e^{-i(\mathcal{E}_k/\hbar)t}|u_k\rangle = e^{-i\omega_k t}|u_k\rangle \tag{3.13}$$

where $\hbar\omega_k = \mathcal{E}_k$ and $|u_k\rangle$ is a solution of the eigenvalue equation:

$$H_0|u_k\rangle = \mathcal{E}_k|u_k\rangle \tag{3.14}$$

If the system is in a state $|\phi_k\rangle$, the average value of a given observable A (corresponding to the operator A) is given by:

$$\langle A\rangle = \langle\phi_k|A|\phi_k\rangle = \langle u_k|A|u_k\rangle, \tag{3.15}$$

which does not change with time: this is why the states $|\phi_k\rangle$ are stationary states for all observables. For example, if the system is in a given stationary state $|\phi_k\rangle$ and the energy is measured, the measurements gives the value $\mathcal{E}_k$:

$$\langle H_0\rangle = \langle\phi_k|H_0|\phi_k\rangle = \langle u_k|H_0|u_k\rangle = \langle u_k|\mathcal{E}_k u_k\rangle = \mathcal{E}_k\langle u_k|u_k\rangle = \mathcal{E}_k. \tag{3.16}$$

A very important general result is that an operator A which corresponds to a physically observable property is Hermitian, that is:

$$\langle\phi|A|\psi\rangle = \langle\psi|A|\phi\rangle^* \tag{3.17}$$

where $|\phi\rangle$ and $|\psi\rangle$ are two given wavevectors. Therefore a Hermitian operator is a self-adjoint operator, since the adjoint operator of A, which is indicated by $A^\dagger$, is given by the following formula:

$$\langle\phi|A|\psi\rangle = \langle\psi|A^\dagger|\phi\rangle^* \tag{3.18}$$

if A is Hermitian $A^\dagger = A$.

It is possible to demonstrate that any two eigenfunctions $|u_n\rangle$ and $|u_m\rangle$ of a Hermitian operator with different eigenvalues, a_n and a_m, are orthogonal:

$$\langle u_n|u_m\rangle = 0 \qquad (m \neq n). \tag{3.19}$$

We will also assume that the eigenfunctions are normalized, so that:

$$\langle u_n|u_m\rangle = \delta_{nm}, \tag{3.20}$$

where δ_{nm} is the Kronecker delta symbol. If the isolated system is not in a particular stationary state, the states $|\phi_k\rangle$ can be used to expand the state function $|\psi\rangle$ of the system:

$$|\psi\rangle = \sum_k c_k|\phi_k\rangle = \sum_k c_k e^{-i\omega_k t}|u_k\rangle, \tag{3.21}$$

where the expansion coefficients c_k do not depend on time, since no external perturbations are applied to the system. These coefficients can be calculated as follows:

$$c_n = \langle\phi_n|\psi\rangle. \tag{3.22}$$

Indeed:

$$\begin{aligned}\langle\phi_n|\psi\rangle &= e^{i\omega_n t}\langle u_n|\psi\rangle = \sum_k e^{i(\omega_n-\omega_k)t}c_k\langle u_n|u_k\rangle = \\ &= \sum_k c_k e^{i(\omega_n-\omega_k)t}\delta_{nk} = c_n\end{aligned} \tag{3.23}$$

where we have used the orthonormality condition 3.20. Given that c_n can be calculated as the scalar product between the generic state $|\psi\rangle$ and the eigenstate $|\phi_n\rangle$, it is useful to think of c_n as the projection of $|\psi\rangle$ on $|u_n\rangle$. Therefore, $|\psi\rangle$ can be represented as a vector in an infinite-dimension abstract space and Eq. 3.21 can be considered as the expansion of a given $|\psi\rangle$ vector in terms of the unit vectors $|u_k\rangle$ in this space.

The physical meaning of the coefficients c_k can be understood by calculating the expectation value of an operator A corresponding to a physical observable:

$$\langle A\rangle = \langle\psi|A|\psi\rangle, \tag{3.24}$$

with $|\psi\rangle$ given by Eq. 3.21. By considering the eigenvalue equation:

$$A|u_n\rangle = a_n|u_n\rangle, \tag{3.25}$$

and Eq. 3.21, $\langle A\rangle$ can be written as:

$$\langle A\rangle = \sum_n a_n|c_n|^2, \tag{3.26}$$

which represents the average of all possible outcomes, a_n, of a measurement of the observable associated to the operator A, weighted by the probability to find the system in a particular eigenstate n, given by $|c_n|^2$.

Exercise 3.1 Calculate the average value of the momentum of an electron in a one-dimensional periodic potential.

The wavefunction describing the system is the Bloch wavefunction:

$$|\psi\rangle = e^{ikx}|u(x)\rangle$$

The momentum operator is:

$$p = -i\hbar\frac{d}{dx}$$

and its average value can be calculated as follows:

$$\langle p\rangle = \langle\psi| - i\hbar\frac{d}{dx}|\psi\rangle$$

where:

$$\langle\psi| = e^{-ikx}\langle u|$$
$$\frac{d|\psi\rangle}{dx} = ike^{ikx}|u\rangle + e^{ikx}\frac{d|u\rangle}{dx}$$

so that:

$$\langle p\rangle = \hbar k\langle u|u\rangle + \langle u| - i\hbar\frac{d}{dx}|u\rangle = \hbar k + \langle p_u\rangle \tag{3.27}$$

where $\langle p_u\rangle$ represents the average momentum of the electron resulting from the interaction with the lattice, since this is the physical meaning of the periodic function $|u(x)\rangle$. Equation 3.27 clearly shows that the electron momentum is composed by two components: the crystal momentum, $\hbar k$, and the momentum due to the interaction with the lattice. ■

3.1.1 Quantum analogy of the classical motion of a particle in confined structures.

Let us consider the simplest superposition of states: we will assume a linear combination of only two states:

$$|\psi_1\rangle = e^{-i\omega_1 t}|u_1(x)\rangle \quad \text{and} \quad |\psi_2\rangle = e^{-i\omega_2 t}|u_2(x)\rangle \tag{3.28}$$

so that:

$$|\psi\rangle = c_1 e^{-i\omega_1 t}|u_1\rangle + c_2 e^{-i\omega_2 t}|u_2\rangle. \tag{3.29}$$

For example we can consider a particle put in a superposition of its ground state $|\psi_1\rangle$, with energy $\mathcal{E}_1 = \hbar\omega_1$, and the first excited state $|\psi_2\rangle$, with energy $\mathcal{E}_2 = \hbar\omega_2$. This superposition state is called wave packet and the variation of the position of the center of mass of this wave packet is the closest quantum mechanical analogy of the classical motion of the particle. The probability density is given by $|\psi(x,t)|^2$,

where:

$$|\psi(x,t)|^2 = \psi(x,t)\psi^*(x,t) = \tag{3.30}$$
$$= |c_1|^2|u_1(x)|^2 + |c_2|^2|u_2(x)|^2 + 2\mathrm{Re}\{c_1^* c_2 u_1^*(x)u_2(x)e^{-i(\omega_2-\omega_1)t}\}.$$

The first two terms do not depend on time, while the third one, generated by the quantum interference between the two states, oscillates in time with a period given by:

$$T = \frac{2\pi}{|\omega_2 - \omega_1|} = \frac{h}{|\mathcal{E}_2 - \mathcal{E}_1|} = \frac{h}{\Delta\mathcal{E}}, \tag{3.31}$$

where $\Delta\mathcal{E} = \mathcal{E}_2 - \mathcal{E}_1$ is the energy separation between the two eigenstates. Therefore, the larger the energy difference $\Delta\mathcal{E}$, the faster is the motion of the particle in the superposition state. On the other hand, the energy separation $\Delta\mathcal{E}$ is related to the mass of the particle and to the spatial extent of the potential well, which confines the particle motion. In this way a link is created between the temporal evolution of the particle dynamics and the spatial extension of the structure where the motion is developing.
If we consider the vibrational levels of a molecule, the typical energy separation is of the order of a few meV, so that the characteristic temporal scale of molecular vibrations ranges from tens to hundreds of femtoseconds (1 fs=10^{-15} s). Indeed, if we take $\Delta\mathcal{E} = 100$ meV we obtain $T \simeq 40$ fs ($h = 4.13 \times 10^{-15}$ eV·s). In the case of valence electrons in atoms, the energy separation can be of the order of a few eV; if we assume $\Delta\mathcal{E} = 10$ eV, we obtain $T \simeq 0.4$ fs. In the case of electrons in core shells $\Delta\mathcal{E} \simeq 1$ keV so that $T \simeq 4$ as (1 as=10^{-18} s). Motion within nuclei is much faster, typically in the zeptosecond (1 zs=10^{-21} s) temporal range. ■

If a perturbation is applied to the system, the Hamiltonian can be written as the sum of the unperturbed Hamiltonian and a perturbation Hamiltonian H':

$$H = H_0 + H' \tag{3.32}$$

This Hamiltonian must be used in the Schrödinger equation 3.10. Also in this case it is useful to expand the state function $|\psi\rangle$ in terms of the eigenfunctions of the unperturbed Hamiltonian and write an expression similar to Eq. 3.21, with the important difference that the coefficients c_k are now time-dependent:

$$|\psi\rangle = \sum_k c_k(t)|\phi_k\rangle = \sum_k c_k(t)e^{-i\omega_k t}|u_k\rangle. \tag{3.33}$$

This expansion can be inserted in the Schrödinger equation 3.10 and a set of differential equations in $c_k(t)$ can be obtained, which can be solved by using various techniques. In the following section we will discuss a particular technique, which can be used to calculate approximated solutions of the Schrödinger equation when a small perturbation is applied to the system. Another technique to study the temporal evolution of a system is based on the density matrix formalism, which we will introduce in Sect. 3.4.

3.2 Time-dependent perturbation theory

Let us consider a system, which at time $t = 0$ is perturbed. The wavefunction $|\psi\rangle$ is the solution of the Schrödinger equation:

$$(H_0 + H')|\psi\rangle = i\hbar\frac{\partial|\psi\rangle}{\partial t}. \tag{3.34}$$

where H' is called interaction Hamiltonian. By using 3.33 in 3.34 we have:

$$\sum_n (H_0 + H')c_n e^{-i\omega_n t}|u_n\rangle = \sum_n \left(i\hbar\frac{dc_n}{dt} + \mathcal{E}_n c_n\right) e^{-i\omega_n t}|u_n\rangle. \tag{3.35}$$

Since $H_0|u_n\rangle = \mathcal{E}_n|u_n\rangle$, Eq. 3.35 can be written as:

$$\sum_n c_n e^{-i\omega_n t} H'|u_n\rangle = i\hbar\sum_n \frac{dc_n}{dt} e^{-i\omega_n t}|u_n\rangle. \tag{3.36}$$

If we now perform the scalar product with $|u_k\rangle$ we obtain:

$$i\hbar\sum_n \frac{dc_n}{dt} e^{-i\omega_n t}\langle u_k|u_n\rangle = \sum_n c_n e^{-i\omega_n t}\langle u_k|H'|u_n\rangle. \tag{3.37}$$

where $\langle u_k|H'|u_n\rangle$ is the matrix element H'_{kn} of H'. By using the orthonormality condition $\langle u_k|u_n\rangle = \delta_{kn}$, the previous equation gives:

$$\frac{dc_k}{dt} = -\frac{i}{\hbar}\sum_n c_n H'_{kn} e^{i\omega_{kn}t}, \tag{3.38}$$

where $\omega_{kn} = \omega_k - \omega_n = (\mathcal{E}_k - \mathcal{E}_n)/\hbar$. So far no approximations have been performed. We will now assume that the Hamiltonian describing the perturbation can be written as $\lambda H'$, where λ is a perturbation parameter, which we will assume very small($\lambda \ll 1$) since we are considering a small perturbation of the system. λ is a very useful parameter since it allows one to distinguish among the various approximation orders in the perturbation calculations. In terms of λ, the power-series expansion for c_n can be written as:

$$c_n = c_n^{(0)} + \lambda c_n^{(1)} + \lambda^2 c_n^{(2)} + \ldots \tag{3.39}$$

By using 3.39 in 3.38 we obtain:

$$\begin{aligned} \frac{dc_k^{(0)}}{dt} &+ \lambda\frac{dc_k^{(1)}}{dt} + \lambda^2\frac{dc_k^{(2)}}{dt} + \ldots = \\ &= -\frac{i}{\hbar}\sum_n [c_n^{(0)} + \lambda c_n^{(1)} + \lambda^2 c_n^{(2)} + \ldots]\lambda H'_{kn} e^{i\omega_{kn}t}. \end{aligned} \tag{3.40}$$

Equating the same powers of λ, the following set of equations is obtained:

$$\begin{aligned} \frac{dc_k^{(0)}}{dt} &= 0 \\ \frac{dc_k^{(1)}}{dt} &= -\frac{i}{\hbar}\sum_n c_n^{(0)} H'_{kn}(t) e^{i\omega_{kn}t} \\ \frac{dc_k^{(2)}}{dt} &= -\frac{i}{\hbar}\sum_n c_n^{(1)} H'_{kn}(t) e^{i\omega_{kn}t}. \end{aligned} \tag{3.41}$$

The solution of the zero-order equation is $c_k^{(0)} = \text{constant}$. Therefore $c_k^{(0)}$ are given by the initial conditions and no transition is occurring. If we assume that at $t = 0$ the system is in the $m-$th state, we can write:

$$\begin{aligned} c_m^{(0)} &= 1 \\ c_n^{(0)} &= 0 \qquad n \neq m \end{aligned} \tag{3.42}$$

so that the second equation in the set 3.41 can be written as:

$$\frac{dc_k^{(1)}}{dt} = -\frac{i}{\hbar} H'_{km} e^{i\omega_{km}t}, \tag{3.43}$$

which gives:

$$c_k^{(1)}(t) = -\frac{i}{\hbar} \int_0^t H'_{km} e^{i\omega_{km}t}\, dt. \tag{3.44}$$

Since at $t = 0$ the system occupies the $m-$th state, $|c_k^{(1)}(t)|^2$ gives the probability (to first order) to find the system in the $k-$th state at time t, therefore it gives the transition probability, $P_{mk}(t)$, from the $m-$state to the $k-$state in the temporal interval t under the action of the external perturbation:

$$P_{mk}(t) = \frac{1}{\hbar^2} \left| \int_0^t H'_{km}(t) e^{i\omega_{km}t}\, dt \right|^2. \tag{3.45}$$

A particularly interesting situation is represented by a harmonic perturbation.

3.2.1 Harmonic perturbation

In the case of a harmonic perturbation with frequency ω, the interaction Hamiltonian can be written as ($t \geq 0$):

$$H'(\mathbf{r}, t) = 2H'(\mathbf{r}) \cos(\omega t) = H'(\mathbf{r}) e^{-i\omega t} + H'^*(\mathbf{r}) e^{i\omega t}, \tag{3.46}$$

and the matrix element H'_{km} is given by:

$$H'_{km} = H'_{km} e^{-i\omega t} + H'^*_{km} e^{i\omega t}. \tag{3.47}$$

Equation 3.44 can be written in the following way:

$$c_k^{(1)} = -\frac{i}{\hbar} \int_0^t [H'_{km} e^{i(\omega_{km}-\omega)t} + H'^*_{km} e^{i(\omega_{km}+\omega)t}]\, dt' \tag{3.48}$$

$$c_k^{(1)} = -\frac{1}{\hbar} \left[H'_{km} \frac{e^{i(\omega_{km}-\omega)t} - 1}{\omega_{km} - \omega} + H'^*_{km} \frac{e^{i(\omega_{km}+\omega)t} - 1}{\omega_{km} + \omega} \right]. \tag{3.49}$$

In the following we will assume that $\omega \approx |\omega_{km}|$: this approximation is called *quasi-resonance approximation* or *rotating phase approximation*. When $\omega \simeq \omega_{km}$ the first term at the right side of Eq. 3.49 is the dominant one, therefore:

$$c_k^{(1)} = -\frac{1}{\hbar} H'_{km} \frac{e^{i(\omega_{km}-\omega)t} - 1}{\omega_{km} - \omega} \tag{3.50}$$

and the transition probability is given by:

$$P_{mk}(t) = |c_k^{(1)}(t)|^2 = \frac{4|H'_{km}|^2}{\hbar^2} \frac{\sin^2\left[\frac{1}{2}(\omega_{km} - \omega)t\right]}{(\omega_{km} - \omega)^2}. \tag{3.51}$$

The first term at the right side of Eq. 3.49 is dominant when $\mathcal{E}_k > \mathcal{E}_m$ and we have an absorption transition: the system, initially in the state m, absorbs a photon with energy $\hbar\omega \approx \mathcal{E}_k - \mathcal{E}_m$ thus producing a transition into a state with higher energy. When $\mathcal{E}_k < \mathcal{E}_m$ and the photon energy of the harmonic perturbation is $\hbar\omega \simeq \mathcal{E}_m - \mathcal{E}_k$, we have a stimulated emission transition from a state with energy $\mathcal{E}_m$ to a state with energy $\mathcal{E}_k$. In this case the transition probability is given by:

$$P_{mk}(t) = \frac{4|H'_{km}|^2}{\hbar^2} \frac{\sin^2\left[\frac{1}{2}(\omega_{km} + \omega)t\right]}{(\omega_{km} + \omega)^2}. \tag{3.52}$$

Figure 3.1 shows the function $f(\Delta\omega) = \sin^2(\Delta\omega t/2)/\Delta\omega^2$, plotted vs. the frequency

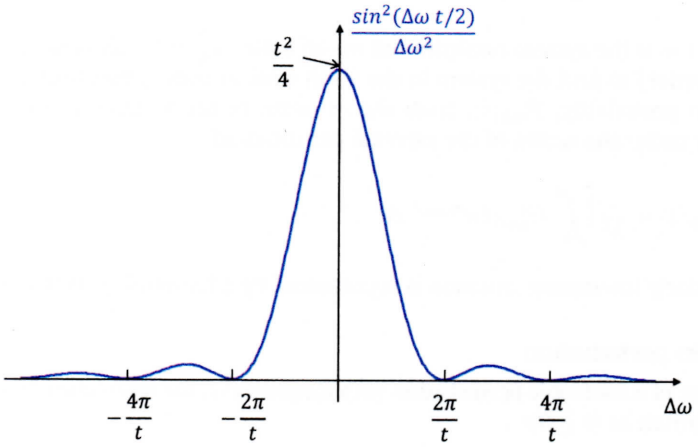

Figure 3.1: Plot of the function $\sin^2(\Delta\omega t/2)/\Delta\omega^2$ vs. $\Delta\omega$.

detuning from resonance, $\Delta\omega = \omega - \omega_{km}$.
When the observation time t is sufficiently long $f(\Delta\omega)$ can be approximated as follows:[1]:

$$\frac{\sin^2\left[\frac{1}{2}(\omega_{km} - \omega)t\right]}{(\omega_{km} - \omega)^2} \simeq \frac{\pi t}{2}\delta(\omega_{km} - \omega), \tag{3.53}$$

where $\delta(\omega_{km} - \omega)$ is the Dirac-δ function and the transition probability can be written as:

$$P_{mk} = \frac{4|H'_{km}|^2}{\hbar^2}\frac{\pi t}{2}\delta(\omega_{km} - \omega) = \frac{2\pi t}{\hbar^2}|H'_{km}|^2\delta(\omega_{km} - \omega). \tag{3.54}$$

The corresponding transition rate is:

$$W_{mk} = \frac{d}{dt}P_{mk}(t) = \frac{2\pi}{\hbar^2}|H'_{km}|^2\delta(\omega_{km} - \omega) \tag{3.55}$$

[1] $\int_{-\infty}^{\infty} \frac{\sin^2 x}{x^2}dx = \pi$, so that $\int_{-\infty}^{\infty} \frac{\sin^2(xt/2)}{x^2}\,dx = \frac{\pi t}{2}$.

Equation 3.55 is called *Fermi's Golden Rule* and it is a particularly important formula. Fermi's golden rule can be written in terms of frequency, ν, and energy $\mathcal{E}$:

$$W_{mk} = \frac{1}{\hbar^2}|H'_{km}|^2\delta(\nu_{km} - \nu) \tag{3.56}$$

$$W_{mk} = \frac{2\pi}{\hbar}|H'_{km}|^2\delta(\mathcal{E}_k - \mathcal{E}_m - \hbar\omega). \tag{3.57}$$

3.2.2 Transitions from a state to a band of states

In several cases of practical importance we have to consider a transition from a state m to a group of states (a band of states) around state k. Let the density of these states be $g(\mathcal{E}_k)$. The transition probability from state m to a final state within this band can be written as:

$$\bar{P}_{mk} = \int_{\text{band},k} P_{mk}(\mathcal{E}_k)\, g(\mathcal{E}_k)\, d\mathcal{E}_k \tag{3.58}$$

where the integral is calculated over the band around energy $\mathcal{E}_k$ and $P_{mk}(\mathcal{E}_k)$ can be obtained from Eq. 3.51:

$$P_{mk}(\mathcal{E}_k) = \frac{|H'_{mk}|^2}{\hbar^2}\frac{\sin^2[t(\mathcal{E}_k - \mathcal{E}_m - \hbar\omega)/2\hbar]}{[(\mathcal{E}_k - \mathcal{E}_m - \hbar\omega)/2\hbar]^2} \tag{3.59}$$

If we now introduce the variable β defined as:

$$\beta = \frac{t}{2\hbar}(\mathcal{E}_k - \mathcal{E}_m - \hbar\omega) \tag{3.60}$$

so that

$$d\beta = \frac{t}{2\hbar}\, d\mathcal{E}_k \tag{3.61}$$

eq. 3.58 can be written in the following way:

$$\bar{P}_{mk} = \frac{2t}{\hbar}\int |H'_{km}|^2 g(\mathcal{E}_k)\frac{\sin^2\beta}{\beta^2}\, d\beta. \tag{3.62}$$

The function $\sin^2\beta/\beta^2$ is very small outside a region around $\beta = 0$ ($\omega = \omega_{km}$) with width $2\pi/t$ (see Fig. 3.1), which can be made arbitrarily small by increasing the observation time t. If $2\pi/t$ is much smaller than the width of the band $g(\omega)$ and we assume that $|H'_{km}|^2$ does not strongly depends on energy, the integral 3.62 becomes:

$$\bar{P}_{mk} = \frac{2t}{\hbar}g(\mathcal{E}_k)|H'_{km}|^2\int_{-\infty}^{\infty}\frac{\sin^2\beta}{\beta^2}d\beta = \frac{2\pi}{\hbar}\, t\, g(\mathcal{E}_k)|H'_{km}|^2. \tag{3.63}$$

The corresponding transition rate is:

$$W_{mk} = \frac{2\pi}{\hbar}|H'_{km}|^2 g(\mathcal{E}_k). \tag{3.64}$$

which can be also written as a function of frequency ν. Indeed, $g(\mathcal{E}_k)\, d\mathcal{E}_k = g(\nu_k)\, d\nu_k$, where $\mathcal{E}_k = h\nu_k$, so that:

$$g(\mathcal{E}_k) = \frac{1}{h}g(\nu_k). \tag{3.65}$$

From 3.64 and 3.65 we find:

$$W_{mk} = \frac{1}{\hbar^2}|H'_{km}|^2 g(\nu_k), \tag{3.66}$$

which represents another form of the Fermi's Golden Rule, which will be used extensively in the next chapters.
As already mentioned, another technique to study a quantum system is based on the use of the density matrix formalism, which is the subject of the final section of this chapter. To introduce the density matrix formalism it is useful to summarise a few general properties of operators, which we have already partially used in this chapter.

3.3 Properties of operators: a short summary

The application of a linear operator A on a ket results in general in another ket, which describes a different state:

$$A|\psi\rangle = |\psi_1\rangle, \tag{3.67}$$

Therefore, an expression such as $\langle\phi|A|\psi\rangle$, which we have already used in this chapter, has the following meaning:

$$\langle\phi|A|\psi\rangle = \langle\phi|\psi_1\rangle, \tag{3.68}$$

that is, it is the inner product of $|\psi_1\rangle$ and $|\phi\rangle$. We have already mentioned the definition of the adjoint operator $A^\dagger$:

$$\langle\phi|A|\psi\rangle = \langle\psi|A^\dagger|\phi\rangle^* \tag{3.69}$$

whose matrix elements are:

$$A^\dagger_{ij} = \langle u_i|A^\dagger|u_j\rangle = \langle u_j|A|u_i\rangle^* = A^*_{ji}. \tag{3.70}$$

If $A = A^\dagger$ the operator A is Hermitian. In this case:

$$A_{ij} = A^*_{ji}, \tag{3.71}$$

therefore the diagonal elements of a Hermitian operator are real ($A_{ii} = A^*_{ii}$).
It is interesting to observe that, while the product between a bra and a ket is a scalar (it is the inner product of the two vectors), the product between a ket and a bra, $|\phi\rangle\langle\psi|$, is an operator. Indeed:

$$\{|\phi\rangle\langle\psi|\}\ |\xi\rangle = |\phi\rangle\langle\psi|\xi\rangle = \langle\psi|\xi\rangle|\phi\rangle \tag{3.72}$$

Since $|\phi\rangle\langle\psi|$ transforms the vector $|\xi\rangle$ in another vector, $\langle\psi|\xi\rangle|\phi\rangle$, it is an operator. This is useful for the definition of the *identity operator*, I, i.e., and operator such that $IA = AI = A$, where A is any operator. By using a set of orthonormal basis vectors, the identity operator can be written as:

$$I = \sum_k |u_k\rangle\langle u_k|. \tag{3.73}$$

Indeed, a generic wavefunction $|\psi\rangle = \sum_k c_k|\phi_k\rangle$, where $c_k = \langle\phi_k|\psi\rangle$, can be written as follows:

$$|\psi\rangle = \sum_k \langle\phi_k|\psi\rangle|\phi_k\rangle = \sum_k |\phi_k\rangle\langle\phi_k|\psi\rangle = \left(\sum_k |u_k\rangle\langle u_k|\right)|\psi\rangle = I|\psi\rangle. \tag{3.74}$$

The matrix elements of I are:

$$I_{ij} = \langle u_i|I|u_j\rangle = \langle u_i|u_j\rangle = \delta_{ij} \tag{3.75}$$

where we have used the fact that $I|u_j\rangle = |u_j\rangle$. The diagonal elements of I are equal to 1 and the off-diagonal ones are equal to 0. We are now ready to introduce the density matrix formalism.

3.4 The density matrix

So far we have considered a system in a *pure state*, described by the wavefunction $|\psi\rangle$. We have seen that the state $|\psi\rangle$ can be written as a superposition of stationary states $|\phi_k\rangle$ and quantum mechanics tells us that the squared modulus of the projection of $|\psi\rangle$ on a given stationary state $|\phi_k\rangle$ gives the probability to find, by a measurement, the system in that particular stationary state. We could confirm this result by preparing the system in a known initial state, perform the measurement (for example one can measure the energy) and repeat several times the same procedure, paying attention to always prepare the system in the same identical initial state. As we have seen, in this case the expectation value of a given operator A can be written as $\langle A\rangle = \sum_n a_n|c_n|^2$, which is nothing but the weighted average of all possible outcomes of the measurement of the physical observable corresponding to the operator A.
Generally we have to deal with statistical (incoherent) mixtures and we cannot know the exact initial quantum state of a system. For example we know only the temperature T of the system, so that the state of the system cannot be described by a wavevector $|\psi\rangle$. What it is possible to know, using for example the laws of statistical mechanics, is the probability, p_n, that the system is in the pure state $|\psi_n\rangle$. Note that the probability p_n is completely different from the probability $|c_n|^2$ discussed before in the case of a pure state $|\psi\rangle$. The probability $|c_n|^2$ does not arise from any lack of knowledge, but it is related to the process of quantum mechanical measurement: unless the system is in a particular eigenstate, the measurement places the system in one of its eigenstates with a probability $|c_n|^2$. On the contrary, the probability p_n results from our lack of information about the exact preparation of the system. In this case the quantum state must be described by a weighted mixture (statistical mixture) of states and it is called a *mixed state*.
For example, it is known that a given macrostate can be described in terms of macroscopic parameters, such as temperature, pressure, volume, etc. A microstate specifies a system in terms of the properties of each of the constituent particles, such as the position and the momentum of each molecule in a gas. From statistical mechanics it is known that many different microstates can correspond to the same macrostate. The probability to find the system in a particular microstate with energy $\mathcal{E}_n$ is:

$$p_n = \frac{\exp(-\mathcal{E}_n/k_BT)}{Z} = \frac{\exp(-\mathcal{E}_n/k_BT)}{\sum_k \exp(-\mathcal{E}_k/k_BT)} \tag{3.76}$$

where Z is the partition function.

We are interested in calculating the average value of an observable corresponding to a given operator A. In the case of a mixed state it is reasonable to write:

$$\langle A\rangle = \sum_n p_n\langle\psi_n|A|\psi_n\rangle, \tag{3.77}$$

indeed, p_n represents the probability that the system is in the pure state $|\psi_n\rangle$ and $\langle\psi_n|A|\psi_n\rangle$ is the average value of A when the system is in this state. Therefore Eq. 3.77 is the probabilistic weighted average of the expectation values of the pure states $|\psi_n\rangle$. It is evident that $\sum_n p_n = 1$. Equation 3.77 can be written in a different way:

$$\langle A\rangle = \sum_n p_n\langle\psi_n|AI|\psi_n\rangle = \sum_{n,k} p_n\langle\psi_n|A|u_k\rangle\langle u_k|\psi_n\rangle, \tag{3.78}$$

where we have used the identity operator. By changing the position of the two scalar products in the previous expression we have:

$$\langle A\rangle = \sum_{n,k} p_n\langle u_k|\psi_n\rangle\langle\psi_n|A|u_k\rangle. \tag{3.79}$$

If we now introduce the *density operator* ρ as:

$$\rho \equiv \sum_n p_n|\psi_n\rangle\langle\psi_n|, \tag{3.80}$$

Eq. 3.79 can be rewritten as:

$$\langle A\rangle = \sum_k \langle u_k|\rho A|u_k\rangle = \mathrm{Tr}(\rho A) \tag{3.81}$$

where $\mathrm{Tr}(\rho A)$ is the trace of the matrix (ρA). Therefore, the expectation value of any operator A is the sum of the diagonal matrix elements of the product ρA. It is evident that, since we can always obtain the average value of an observable as $\mathrm{Tr}(\rho A)$, the density operator ρ must contain all the physically relevant information that we can know about a system. Note that in the definition of ρ (Eq. 3.80), the state functions $|\psi_n\rangle$ are normalized but are not necessarily orthogonal.

An important result is that ρ is Hermitian. Indeed:

$$\begin{aligned}\langle\psi|\rho|\phi\rangle &= \sum_n p_n\langle\psi|\psi_n\rangle\langle\psi_n|\phi\rangle = \sum_n p_n\langle\psi_n|\psi\rangle^*\langle\phi|\psi_n\rangle^* = \\ &= \sum_n p_n(\langle\phi|\psi_n\rangle\langle\psi_n|\psi\rangle)^* = \langle\phi|\rho|\psi\rangle^*\end{aligned} \tag{3.82}$$

Moreover $\mathrm{Tr}(\rho) = 1$:

$$\begin{aligned}\mathrm{Tr}(\rho) &= \sum_k \langle u_k|\rho|u_k\rangle = \sum_{k,n} p_n\langle u_k|\psi_n\rangle\langle\psi_n|u_k\rangle = \\ &= \sum_{k,n} p_n\langle\psi_n|u_k\rangle\langle u_k|\psi_n\rangle = \sum_n p_n\langle\psi_n|\psi_n\rangle = 1.\end{aligned} \tag{3.83}$$

By writing the pure states $|\psi_n\rangle$ in 3.80 as:

$$|\psi_n\rangle = \sum_k c_{nk}e^{-i\omega_k t}|u_k\rangle$$

the density operator can be written as

$$\rho = \sum_{n,k,\ell} p_n c_{nk}c^*_{n\ell}e^{-i(\omega_k-\omega_\ell)t}|u_k\rangle\langle u_\ell|. \tag{3.84}$$

The matrix elements ρ_{ij} are given by:

$$\begin{aligned}\rho_{ij} &= \langle u_i|\rho|u_j\rangle = \sum_{n,k,\ell} p_n c_{nk} c^*_{n\ell} e^{-i(\omega_k-\omega_\ell)t}\langle u_i|u_k\rangle\langle u_\ell|u_j\rangle = \\ &= \sum_n p_n c_{ni} c^*_{nj} e^{-i(\omega_i-\omega_j)t} \end{aligned} \tag{3.85}$$

the diagonal elements are:

$$\rho_{jj} = \sum_n p_n |c_{nj}|^2. \tag{3.86}$$

so that ρ_{jj} gives the probability to find the system in the state $|u_j\rangle$. The off-diagonal elements are called *coherence elements* and their physical meaning will be discussed in Sect. 4.5.1.

Exercise 3.2 Consider a pure state $|\psi\rangle$. Calculate $\mathrm{Tr}(\rho)$, $\mathrm{Tr}(\rho^2)$, the expectation value of the observable associated to the operator A, the diagonal and the off-diagonal elements of ρ.

The density operator is:

$$\rho = |\psi\rangle\langle\psi| \tag{3.87}$$

Its trace can be easily calculated:

$$\mathrm{Tr}(\rho) = \sum_k \langle u_k|\rho|u_k\rangle = \sum_k \langle u_k|\psi\rangle\langle\psi|u_k\rangle = \sum_k \langle\psi|u_k\rangle\langle u_k|\psi\rangle = \langle\psi|\psi\rangle = 1. \tag{3.88}$$

Now we calculate ρ^2:

$$\rho^2 = (|\psi\rangle\langle\psi|)\,(|\psi\rangle\langle\psi|) = |\psi\rangle\langle\psi|\psi\rangle\langle\psi| = |\psi\rangle\langle\psi| = \rho \tag{3.89}$$

Therefore

$$\mathrm{Tr}(\rho^2) = \mathrm{Tr}(\rho) = 1. \tag{3.90}$$

The expectation value of A is:

$$\begin{aligned}\langle A\rangle &= \mathrm{Tr}(\rho A) = \sum_k \langle u_k|\rho A|u_k\rangle = \sum_k \langle u_k|\psi\rangle\langle\psi|A|u_k\rangle = \\ &= \sum_k a_k \langle u_k|\psi\rangle\langle\psi|u_k\rangle = \sum_k a_k |c_k|^2 \end{aligned} \tag{3.91}$$

where we have used the eigenvalue equation $A|u_k\rangle = a_k|u_k\rangle$.
The diagonal elements of ρ are:

$$\begin{aligned}\rho_{jj} &= \langle u_j|\rho|u_j\rangle = \langle u_j|\psi\rangle\langle\psi|u_j\rangle = \langle u_j|\psi\rangle\langle u_j|\psi\rangle^* = \\ &= c_j e^{-i\omega_j t} c_j^* e^{i\omega_j t} = |c_j|^2. \end{aligned} \tag{3.92}$$

The off-diagonal elements of ρ are:

$$\begin{aligned}\rho_{ij} &= \langle u_i|\rho|u_j\rangle = \langle u_i|\psi\rangle\langle\psi|u_j\rangle = \langle u_i|\psi\rangle\langle u_j|\psi\rangle^* = \\ &= c_i c_j^* e^{-i(\omega_i-\omega_j)t}. \end{aligned} \tag{3.93}$$

In the absence of relaxation processes the populations of the different energy levels, given by the diagonal elements of ρ, are time inariant so that the time dependence is present only in the off-diagomal matrix elements. ■

3.4.1 Density matrix of a canonical ensemble.

A *canonical ensemble* is a statistical ensemble of systems which are all in contact (i.e., in thermal equilibrium) with a heat reservoir at temperature T, so that the overall system is in thermal equilibrium with its environment. The corresponding density matrix can be written as:

$$\rho = \frac{e^{-\beta H}}{Z}, \tag{3.94}$$

where H is the Hamiltonian operator, $\beta = 1/k_B T$ and Z is the canonical partition function defined in Eq. 3.76, which can be written as:

$$Z = \mathrm{Tr}(e^{-\beta H}) = \sum_n e^{-\beta \mathcal{E}_n}. \tag{3.95}$$

Since we have (if $H|u_k\rangle = \mathcal{E}_k|u_k\rangle$):

$$e^{-\beta H}|u_k\rangle = e^{-\beta \mathcal{E}_k}|u_k\rangle, \tag{3.96}$$

the matrix element of the density matrix can be calculated as:

$$\rho_{ij} = \langle u_i|\frac{e^{-\beta H}}{Z}|u_j\rangle = \frac{e^{-\beta \mathcal{E}_j}}{Z}\delta_{ij}, \tag{3.97}$$

thus meaning that the density matrix is diagonal with diagonal elements given by $e^{-\beta \mathcal{E}_j}/Z$. Therefore if the system is initially out of thermal equilibrium, it must evolve so that the diagonal elements redistribute to a Boltzmann distribution and the off-diagonal elements must go to zero. In the first case we have a population relaxation, in the second case we have a dephasing process. The relaxation processes will be discussed in Sect. 4.5.1. ■

For a **mixed state** the property 3.89 does not hold. Indeed in general:

$$\rho^2 = \sum_n \sum_m p_n p_m |\psi_n\rangle\langle\psi_n|\psi_m\rangle\langle\psi_m| \neq \rho, \tag{3.98}$$

moreover, it is possible to demonstrate that:

$$\mathrm{Tr}(\rho^2) \leq 1. \tag{3.99}$$

In Eq. 3.99 $\mathrm{Tr}(\rho^2) = 1$ if and only if the system can be in only one physical state, i.e., all but one of the probabilities p_n corresponding to independent states must be zero. It is possible to conclude that a state of a physical system is a pure state if $\mathrm{Tr}(\rho^2) = 1$, whereas if $\mathrm{Tr}(\rho^2) < 1$ the system is in a mixed state. $\mathrm{Tr}(\rho^2)$ can be interpreted as a measure of the *mixed nature* of a given system.

Exercise 3.3 Demonstrate that in general $\mathrm{Tr}(\rho^2) \leq 1$.

$$\mathrm{Tr}(\rho^2) = \sum_k \langle u_k| \left\{ \sum_{nm} p_n p_m |\psi_n\rangle\langle\psi_n|\psi_m\rangle\langle\psi_m| \right\} |u_k\rangle =$$
$$= \sum_k \sum_{nm} p_n p_m \langle u_k|\psi_n\rangle\langle\psi_n|\psi_m\rangle\langle\psi_m|u_k\rangle =$$
$$= \sum_k \sum_{nm} p_n p_m \langle\psi_m|u_k\rangle\langle u_k|\psi_n\rangle\langle\psi_n|\psi_m\rangle \tag{3.100}$$

in the last expression we recognize the identity operator I, so that:

$$\mathrm{Tr}(\rho^2) = \sum_{nm} p_n p_m \langle\psi_m|\psi_n\rangle\langle\psi_n|\psi_m\rangle = \sum_{nm} p_n p_m |\langle\psi_n|\psi_m\rangle|^2 \tag{3.101}$$

From the Schwarz inequality one has:

$$|\langle\psi_n|\psi_m\rangle|^2 \leq \langle\psi_n|\psi_n\rangle\langle\psi_m|\psi_m\rangle \tag{3.102}$$

so that:

$$\mathrm{Tr}(\rho^2) \leq \sum_{nm} p_n p_m \langle\psi_n|\psi_n\rangle\langle\psi_m|\psi_m\rangle = \sum_n p_n \sum_m p_m = 1. \tag{3.103}$$

■

3.4.1 Temporal dependence of ρ

We have seen that by using the density matrix formalism it is possible to obtain the average value of a given observable. What we need to know in addition is the temporal evolution of ρ. We start from the definition of ρ:

$$\rho = \sum_n p_n |\psi_n\rangle\langle\psi_n|,$$

by calculating the temporal derivative we obtain:

$$\frac{\partial\rho}{\partial t} = \sum_n p_n \left[\frac{\partial|\psi_n\rangle}{\partial t}\langle\psi_n| + |\psi_n\rangle\frac{\partial\langle\psi_n|}{\partial t} \right] \tag{3.104}$$

The Schrödinger equation for $|\psi_n\rangle$ is

$$H|\psi_n\rangle = i\hbar\frac{\partial|\psi_n\rangle}{\partial t}, \tag{3.105}$$

the corresponding equation for $\langle\psi_n|$ is:

$$\langle\psi_n|H = -i\hbar\frac{\partial\langle\psi_n|}{\partial t} \tag{3.106}$$

By using Eqs. 3.105 and 3.106 in 3.104 we obtain:

$$i\hbar\frac{\partial\rho}{\partial t} = \sum_n p_n [H|\psi_n\rangle\langle\psi_n| - |\psi_n\rangle\langle\psi_n|H] = [H\rho - \rho H]. \tag{3.107}$$

The operator $H\rho - \rho H$ is called *commutator* and can be written as $[H, \rho]$. Therefore we have:

$$i\hbar\frac{\partial\rho}{\partial t} = [H, \rho] \tag{3.108}$$

or, equivalently:

$$\frac{\partial\rho}{\partial t} = \frac{i}{\hbar}[\rho, H]. \tag{3.109}$$

called the *von Neumann equation*, which is the quantum analog of the classical Liouville equation. In Section 4.5 this equation will be used to calculate the susceptibility of a two-level system, which we will subsequently extend to the case of a semiconductor.

4. Electron-photon interaction

4.1 Introduction

In this section we will study the interaction of a quantized medium with an electromagnetic field, which will be treated classically. This approach is called *semiclassical approximation*. After an introduction on general aspects of electromagnetic theory, which will be used in this chapter, the Hamiltonian describing the interaction of electrons with an electromagnetic field will be introduced. Finally, by using the density matrix formalism the linear optical susceptibility of a two-level system will be calculated.

4.2 Classical electromagnetic theory

Electric and magnetic fields satisfy Maxwell's equations:

$$\begin{aligned}
&\nabla \cdot \mathbf{D} = \rho \\
&\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \\
&\nabla \cdot \mathbf{B} = 0 \\
&\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}
\end{aligned} \tag{4.1}$$

where $\mathbf{E}$ and $\mathbf{B}$ are the electric and magnetic fields, $\mathbf{D} = \epsilon \mathbf{E}$ is the electric displacement vector, ϵ is the dielectric constant given by $\epsilon = \epsilon_r \epsilon_0$, $\mathbf{H} = \mathbf{B}/\mu$, where μ is the permeability (we will assume $\mu = \mu_0$), ρ is the carrier density and $\mathbf{J}$ the current density. Since $\nabla \cdot \mathbf{B} = 0$, the magnetic field can be written as the curl of a vector potential:

$$\mathbf{B} = \nabla \times \mathbf{A}. \tag{4.2}$$

This equation does not define $\mathbf{A}$ in a unique way, indeed we can define a new vector potential $\mathbf{A}'$ as:

$$\mathbf{A}' = \mathbf{A} + \nabla f, \tag{4.3}$$

where f is an arbitrary scalar function. Indeed, from Eq. 4.3:

$$\nabla \times \mathbf{A}' = \nabla \times \mathbf{A} + \nabla \times (\nabla f) = \nabla \times \mathbf{A}. \tag{4.4}$$

Since we have defined the vector potential in terms of its curl such that $\mathbf{B} = \nabla \times \mathbf{A}$, its divergence is arbitrary. The choice of the value of $\nabla \cdot \mathbf{A}$ defines the particular gauge within which to study a particular process. For the investigation of electron-photon interactions it is common practice to adopt the so-called *radiation or Coulomb gauge*, assuming that $\nabla \cdot \mathbf{A} = 0$. If we calculate the divergence of the vectors in Eq. 4.3 we obtain:

$$\nabla \cdot \mathbf{A}' = \nabla \cdot \mathbf{A} + \nabla \cdot \nabla f = \nabla \cdot \mathbf{A} + \nabla^2 f, \tag{4.5}$$

it is always possible to choose a particular function f in such a way that $\nabla^2 f + \nabla \cdot \mathbf{A} = 0$. Within this gauge the solutions for the vector potential are represented by transverse electromagnetic waves. From Maxwell's equations we have:

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} = -\frac{\partial}{\partial t}(\nabla \times \mathbf{A}) = -\nabla \times \left(\frac{\partial \mathbf{A}}{\partial t}\right), \tag{4.6}$$

from which we obtain:

$$\nabla \times \left(\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t}\right) = 0. \tag{4.7}$$

Since the vectorial field $\mathbf{E} + \partial \mathbf{A}/\partial t$ is characterized by a curl equal to zero, it is conservative and can be written as the gradient of a scalar potential ϕ:

$$\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t} = -\nabla \phi, \qquad \mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t} - \nabla \phi. \tag{4.8}$$

Also Eq. 4.8 does not define ϕ in a unique way. If the vector potential is written as in 4.3, the scalar potential can be written as:

$$\phi' = \phi - \frac{\partial f}{\partial t}. \tag{4.9}$$

without producing any effect on the electric and magnetic fields. In a region where $\mathbf{J} = 0$ and $\rho = 0$ we can assume a constant background potential $\phi = 0$ and the fields are given by:

$$\mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t}, \qquad \mathbf{B} = \nabla \times \mathbf{A}. \tag{4.10}$$

In a region free from carriers and currents the vector potential satisfies the homogeneous wave equation:

$$\nabla^2 \mathbf{A} - \frac{1}{c^2}\frac{\partial^2 \mathbf{A}}{\partial t^2} = 0, \tag{4.11}$$

a plane wave solution of the previous equation is given by:

$$\mathbf{A}(\mathbf{r}, t) = \hat{\epsilon} A_0 \sin(\mathbf{k} \cdot \mathbf{r} - \omega t), \tag{4.12}$$

where $\hat{\epsilon}$ is the polarization unit vector, $\mathbf{k}$ is the propagation vector of the electromagnetic field, ω is the angular frequency, which is related to $\mathbf{k}$ by the following relation:

$$\omega = kc. \tag{4.13}$$

The Coulomb gauge $\nabla \cdot \mathbf{A} = 0$ is satisfied if:

$$\mathbf{k} \cdot \hat{\epsilon} = 0 \tag{4.14}$$

i.e., $\mathbf{k}$ is perpendicular to the polarization direction, so that the wave is transverse. The corresponding electric field is:

$$\mathbf{E}(\mathbf{r}, t) = \hat{\epsilon}\omega A_0 \cos(\mathbf{k} \cdot \mathbf{r} - \omega t) = \hat{\epsilon} E_0 \cos(\mathbf{k} \cdot \mathbf{r} - \omega t). \tag{4.15}$$

Since $\mathbf{B} = \nabla \times \mathbf{A}$ and $\omega = kc$, the magnetic field can be written as:

$$\mathbf{B}(\mathbf{r}, t) = \frac{E_0}{c}\left(\frac{\mathbf{k}}{k} \times \hat{\epsilon}\right) \cos(\mathbf{k} \cdot \mathbf{r} - \omega t) = \frac{E_0}{c}(\hat{\mathbf{k}} \times \hat{\epsilon}) \cos(\mathbf{k} \cdot \mathbf{r} - \omega t), \tag{4.16}$$

where $\hat{\mathbf{k}} = \mathbf{k}/k$ is the unit propagation vector. $\mathbf{k}$, $\mathbf{E}$ and $\mathbf{B}$ are mutually perpendicular and $|\mathbf{E}|/|\mathbf{B}| = c$.
The energy density of an electromagnetic field is given by:

$$u(\mathbf{r}, t) = \frac{1}{2}\left\{\epsilon_0[\mathbf{E}(\mathbf{r}, t)]^2 + \frac{1}{\mu_0}[\mathbf{B}(\mathbf{r}, t)]^2\right\}. \tag{4.17}$$

The rate of energy flow through a unit area normal to the propagation direction is given by the Poynting vector:

$$\mathbf{S} = \frac{1}{\mu_0}(\mathbf{E} \times \mathbf{B}). \tag{4.18}$$

Using the expressions for th electric (Eq. 4.15) and magnetic (Eq. 4.16) fields in Eqs. 4.17 and 4.18 we have:

$$u(\mathbf{r}, t) = \epsilon_0[E_0 \cos(\mathbf{k} \cdot \mathbf{r} - \omega t)]^2 \tag{4.19}$$

and

$$\mathbf{S}(\mathbf{r}, t) = \epsilon_0 c[E_0 \cos(\mathbf{k} \cdot \mathbf{r} - \omega t)]^2 \hat{\mathbf{k}} \tag{4.20}$$

so that

$$\mathbf{S}(\mathbf{r}, t) = c\, u(\mathbf{r}, t)\hat{\mathbf{k}}. \tag{4.21}$$

If we now average the energy density and the Poynting vector over a period $T = 2\pi/\omega$ we obtain:

$$u = \frac{1}{2}\epsilon_0 E_0^2 = \frac{1}{2}\epsilon_0 \omega^2 A_0^2 \tag{4.22}$$

and

$$\langle \mathbf{S}(t) \rangle_t = \frac{1}{2}\epsilon_0 c E_0^2 \hat{\mathbf{k}} \tag{4.23}$$

and the intensity of the electromagnetic field is:

$$I = \frac{1}{2}\epsilon_0 c E_0^2. \tag{4.24}$$

It is now simple to find a relation between the amplitude of the electric field and the photon flux associated to an electromagnetic wave. If $N(\omega)$ is the number of photons with energy $\hbar\omega$ within a volume V, the energy density can be written as:

$$u = \frac{\hbar\omega N(\omega)}{V}. \tag{4.25}$$

By using Eqs. 4.22 and 4.25 we obtain:

$$E_0^2 = \frac{2\hbar\omega N(\omega)}{\epsilon_0 V}, \qquad A_0^2 = \frac{2\hbar N(\omega)}{\epsilon_0 \omega V}. \tag{4.26}$$

After this brief summary of the properties of electromagnetic field, we can analyse the interaction of an electron with such a field.

4.3 Electrons in an electromagnetic field

The non-relativistic Hamiltonian of a particle with mass m and charge q in an electromagnetic field described by the vector potential $\mathbf{A}(\mathbf{r},t)$ and the scalar potential $\phi(\mathbf{r},t)$ is:

$$H = \frac{[\mathbf{p} - q\mathbf{A}(\mathbf{r},t)]^2}{2m} + V(\mathbf{r}) + q\phi(\mathbf{r},t) \tag{4.27}$$

where $V(\mathbf{r})$ is some effective potential. In the previous expression $\mathbf{p}$ is the *canonical momentum* of the charged particle, which is related to the usual particle momentum $m\mathbf{v}$ by the following equation:

$$\mathbf{p} = m\mathbf{v} + q\mathbf{A}. \tag{4.28}$$

We note that the Hamiltonian 4.27 is not unique, but it is just one of the possible Hamiltonians, which can be obtained by starting from a suitable Lagrangian $\mathcal{L}(q,\dot{q},t)$, where q are generalised coordinates and the dot over a variable indicates a total derivative with respect to time. In the case of a conservative system the Lagrangian is defined as the difference between the kinetic and the potential energies. Since the electromagnetic field is not conservative a generalised definition is required. We assume that $\mathcal{L}$ is a suitable Lagrangian if, when inserted in the Lagrange's equations:

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{q}_i}\right) - \frac{\partial \mathcal{L}}{\partial q_i} = 0, \tag{4.29}$$

it gives the correct equation of motion of the particle of mass m, charge q and velocity $\dot{\mathbf{r}}$ in an electromagnetic field and in an effective potential $V(\mathbf{r})$. A suitable Lagrangian is the following:

$$\mathcal{L} = \frac{1}{2}m\dot{\mathbf{r}}^2 - V(\mathbf{r}) - q\phi + q\dot{\mathbf{r}}\cdot\mathbf{A}. \tag{4.30}$$

In Cartesian coordinates (i.e., $q_i = x_i$, with $i = 1,2,3$) the previous expression can be written as:

$$\mathcal{L} = \frac{1}{2}m\sum_i \dot{x}_i^2 - V(\mathbf{r}) - q\phi + q\sum_i \dot{x}_i A_i. \tag{4.31}$$

The components of the generalised canonical momentum are calculated by differentiating the Lagrangian with respect to the generalised velocities:

$$p_i = \frac{\partial \mathcal{L}}{\partial \dot{q}_i} \tag{4.32}$$

By using the Lagrangian 4.30, p_i are given by:

$$p_i = m\dot{x}_i + qA_i \tag{4.33}$$

so that the canonical momentum is given by Eq. 4.28. The Hamiltonian associated to the Lagrangian 4.30 can be calculated by employing the following Legendre transformation:

$$H(q,p,t) = \sum_i \dot{q}_i p_i - \mathcal{L}(q,\dot{q},t) \tag{4.34}$$

and it is given by 4.27.

Exercise 4.1 Demonstrate that the Hamiltonian 4.27 can be obtained from the Lagrangian 4.30

The Hamiltonian can be calculated by using the transformation 4.34, which, in Cartesian coordinate can be written as:

$$H = \sum_i \dot{x}_i p_i - \mathcal{L}, \tag{4.35}$$

where the generalised velocities can be obtained from Eq. 4.33:

$$\dot{x}_i = \frac{p_i - qA_i}{m}. \tag{4.36}$$

If we now use this expression in Eq. 4.35 we obtain:

$$H = \sum_i \frac{(p_i - qA_i)^2}{2m} + V + q\phi. \tag{4.37}$$

■

If we also consider the energy of the electromagnetic field, the total Hamiltonian can be written as:

$$H_t = H + H_{field} \tag{4.38}$$

where H is given by Eq. 4.27 and

$$H_{field} = \frac{1}{2}\epsilon_0 \int (E^2 + c^2 B^2)\, d\tau. \tag{4.39}$$

In several cases of practical interest the electromagnetic energy is much larger than the interaction energy between field and charged particles. Therefore, we can completely neglect the variation of the field energy during the interaction. In this case the fields are virtually not influenced by the interaction and can be treated in a classical way (i.e., we do not have to quantize the field). This is at the heart of the semi-classical approximation. In this case $H_t = H$ and can be written as:

$$H = H_0 + H' \tag{4.40}$$

where:

$$H_0 = \frac{\mathbf{p}^2}{2m} + V(\mathbf{r}) = -\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r}) \tag{4.41}$$

is the unperturbed Hamiltonian and

$$\begin{aligned} H' &= -\frac{q}{2m}(\mathbf{p}\cdot\mathbf{A} + \mathbf{A}\cdot\mathbf{p}) + \frac{q^2A^2}{2m} + q\phi = \\ &= i\hbar\frac{q}{2m}(\nabla\cdot\mathbf{A} + \mathbf{A}\cdot\nabla) + \frac{q^2A^2}{2m} + q\phi \end{aligned} \tag{4.42}$$

is the interaction Hamiltonian. As already pointed out, we can assume $\phi = 0$, moreover for many applications we can neglect the term $q^2A^2/2m$ with respect to the term $(q/2m)\mathbf{p}\cdot\mathbf{A}$. This approximation is valid when the intensity of the electromagnetic field is not too high. As shown in Ex. 4.2, even at an intensity of 1 MW/cm^2 it is still possible to neglect the term proportional to A^2 since $|\mathbf{p}| \gg |e\mathbf{A}|$ (*weak-field limit*).

Exercise 4.2 Consider an electromagnetic field with intensity $I = 10^6$ W/cm^2 and a photon energy of 1 eV. Assuming an electron velocity of 5×10^6 cm/s, calculate the ratio eA/p.

The amplitude of the electric field amplitude can be calculated by using Eq. 4.24:

$$E_0 = \sqrt{2I/\epsilon_0 c} = 2.7 \times 10^6 \text{ V/m},$$

In the case of a photon energy of 1 eV (angular frequency $\omega = 1.52 \times 10^{15}$ rad/s) the amplitude of the vector potential is:

$$A_0 = \frac{E_0}{\omega} = 1.8 \times 10^{-9} \text{ Vs/m}.$$

If we now calculate the ratio between the terms $q^2A^2/2m$ and $qpA/2m$ in the case of electrons ($q = -e = -1.6 \times 10^{-19}$ C), assuming an electron velocity of 5×10^6 cm/s, we obtain:

$$eA/p \simeq 6 \times 10^{-3}.$$

■

From the physical point of view the term proportional to $\mathbf{A}$ is related to the transitions involving the emission or the absorption of a single photon, while the term proportional to A^2 is related to the transitions involving the emission or the absorption of two photons, which, in the case of low excitation intensity, are much less probable. Note that the operators $\mathbf{p} = -i\hbar\nabla$ and $\mathbf{A}$ commute. Indeed, in the Coulomb gauge ($\nabla\cdot\mathbf{A} = 0$) we have:

$$\nabla\cdot(\mathbf{A}|\psi\rangle) = (\nabla\cdot\mathbf{A})|\psi\rangle + \mathbf{A}\cdot\nabla|\psi\rangle = \mathbf{A}\cdot\nabla|\psi\rangle \tag{4.43}$$

so that

$$\nabla\cdot\mathbf{A} = \mathbf{A}\cdot\nabla \tag{4.44}$$

and the commutator $[-i\hbar\nabla, \mathbf{A}] = 0$. The interaction Hamiltonian given by Eq. 4.42 can be written as:

$$H' = -\frac{q}{m}\,\mathbf{A}\cdot\mathbf{p} \tag{4.45}$$

In the case of electrons $q = -e$ and

$$H' = \frac{e}{m}\,\mathbf{A}\cdot\mathbf{p} \tag{4.46}$$

4.4 Electric dipole approximation

Let us assume that the wavelength λ of the electromagnetic field is large compared to the size of the atomic system under consideration and that the field intensity is not too high. When these two conditions are satisfied, the *dipole approximation* can be used. Within this approximation the spatial variation of the electromagnetic field across the atom can be neglected. This situation is usually met since the size of typical atomic systems is of the order of $0.1 - 10$ nm, much smaller than the wavelength of electromagnetic radiation of interest in photonic and electronic applications (typically in the ultraviolet-visible spectral regions). Within this approximation, if we consider an atom whose nucleus is located at the position $\mathbf{R}$, the vector potential can be assumed as spatially homogeneous so that:

$$\mathbf{A}(\mathbf{r}, t) \approx \mathbf{A}(\mathbf{R}, t) \tag{4.47}$$

This means that we can neglect the spatial variation of $\mathbf{A}$ over atomic dimensions. In this case the Lagrangian given by Eq. 4.30 can be written as:

$$\mathcal{L} = \frac{1}{2}m\dot{\mathbf{r}}^2 - V(\mathbf{r}) - q\phi + q\dot{\mathbf{r}}\cdot\mathbf{A}(\mathbf{R}, t). \tag{4.48}$$

It is a general result of calculus that if a total time derivative is added to a Lagrangian, the equations of motion derived from it do not change (in other words, two Lagrangian which differ by a total time derivative, $\mathcal{L}' = \mathcal{L} + \frac{d}{dt}F(q, t)$, are equivalent). Therefore we may add to the Lagrangian 4.48 the following total time derivative:

$$-q\frac{d}{dt}[\mathbf{r}\cdot\mathbf{A}(\mathbf{R}, t)], \tag{4.49}$$

in this way an equivalent Lagrangian is obtained:

$$\mathcal{L} = \frac{1}{2}m\dot{\mathbf{r}}^2 - V(\mathbf{r}) - q\phi - q\mathbf{r}\cdot\dot{\mathbf{A}}(\mathbf{R}, t). \tag{4.50}$$

The canonical momentum $p_i = \partial\mathcal{L}/\partial\dot{q}_i$, is given by:

$$\mathbf{p} = m\dot{\mathbf{r}} = m\frac{d\mathbf{r}}{dt} \tag{4.51}$$

and the corresponding Hamiltonian, which can be calculated by using Eq. 4.34, can be written as follows:

$$H = \frac{\mathbf{p}^2}{2m} + V(\mathbf{r}) + q\phi + q\mathbf{r}\cdot\frac{d\mathbf{A}(\mathbf{R}, t)}{dt}. \tag{4.52}$$

In a region free from charged particles we can assume $\phi = 0$ and Eq. 4.52 can be rewritten as:

$$H = \frac{\mathbf{p}^2}{2m} + V(\mathbf{r}) - \boldsymbol{\mu} \cdot \mathbf{E}(\mathbf{R}, t), \tag{4.53}$$

where:

$$\boldsymbol{\mu} = q\mathbf{r} = -e\mathbf{r} \tag{4.54}$$

is the *electric dipole moment* and we have used the fact that, over atomic dimensions:

$$\frac{d\mathbf{A}(\mathbf{R}, t)}{dt} = -\mathbf{E}(\mathbf{R}, t). \tag{4.55}$$

Therefore, in the framework of the electric dipole approximation, the interaction Hamiltonian is given by:

$$H' = -\boldsymbol{\mu} \cdot \mathbf{E}. \tag{4.56}$$

Note that the electric dipole approximation works well even in the case of electromagnetic fields extending into the far ultraviolet.
The matrix elements of the interaction Hamiltonian H' given by Eq. 4.56 are:

$$H'_{mn} = \langle u_m | H' | u_n \rangle = -\langle u_m | \boldsymbol{\mu} \cdot \mathbf{E} | u_n \rangle, \tag{4.57}$$

where $|u_i\rangle$ are the eigenvectors of the unperturbed Hamiltonian H_0. Since in the electric dipole approximation the electric field is a function only of time, in Eq. 4.57 $\mathbf{E}$ can be taken outside the scalar product. Therefore:

$$H'_{mn} = -\langle u_m | \boldsymbol{\mu} | u_n \rangle \cdot \mathbf{E} = -\boldsymbol{\mu}_{mn} \cdot \mathbf{E}. \tag{4.58}$$

If $H_0(\mathbf{r}) = H_0(-\mathbf{r})$, i.e., if the Hamiltonian is invariant under the transformation $\mathbf{r} \to -\mathbf{r}$:

$$|u_k(-\mathbf{r})\rangle = \pm |u_k(\mathbf{r})\rangle, \tag{4.59}$$

i.e., the eigenstates have definite parity. When the eigenfunction is an even function of $\mathbf{r}$ it is called eigenfunction of even parity and the associated state is called state with even parity. If $|u_k(-\mathbf{r})\rangle = -|u_k(\mathbf{r})\rangle$ the wavefunction has odd parity and the associated state is called state with odd parity. For example, $H_0(\mathbf{r}) = H_0(-\mathbf{r})$ when the system is characterized by a center of symmetry, or when the potential energy $V(\mathbf{r})$ is a function only of the distance between particles. Therefore, if two eigenstates $|u_m\rangle$ and $|u_n\rangle$ with different eigenvalues, $\mathcal{E}_m$ and $\mathcal{E}_n$, have the same parity, the matrix elements μ_{mn} are zero since in the integral $\int u_m^*(\mathbf{r})\ \mathbf{r}\ u_n(\mathbf{r})\ d\tau = \langle u_m | \mathbf{r} | u_n \rangle$ the integrand is an odd function. The consequence is that the transition between states with the same parity, which is directly related to the matrix element H'_{mn}, is forbidden in the electric dipole approximation. The transition can be possible but in higher-order processes, for example as the electric quadrupole transition, which however is much less probable. Therefore we can conclude that electric dipole transitions can occur only between states with opposite parity. If the eigenstates $|u_m\rangle$ and $|u_n\rangle$ have definite parity and are non-degenerate, the diagonal matrix elements of $\boldsymbol{\mu}$ are zero.

In the following section we will apply the density matrix formalism to calculate the linear optical susceptibility in the case of a two-level system excited by a plane monochromatic electromagnetic radiation, in the framework of the dipole approximation. Then, in the next chapter, we will generalize the results to the case of a semiconductor.

4.5 Linear optical susceptibility

Let us consider a two-level system characterized by the unperturbed Hamiltonian H_0. We will consider two eigenstates $|u_1\rangle$ and $|u_2\rangle$ with energies $\mathcal{E}_1$ and $\mathcal{E}_2$, respectively: $H_0|u_i\rangle = \mathcal{E}_i|u_i\rangle$ $(i = 1, 2)$. The matrix elements of H_0 are the following:

$$H_{0,mn} = \langle u_m|H_0|u_n\rangle = \langle u_m|\mathcal{E}_n u_n\rangle = \mathcal{E}_n\langle u_m|u_n\rangle = \mathcal{E}_n\delta_{mn} \tag{4.60}$$

so that

$$H_0 = \begin{pmatrix} \mathcal{E}_1 & 0 \\ 0 & \mathcal{E}_2 \end{pmatrix} \tag{4.61}$$

We assume that at time $t = 0$ a perturbation is applied to the system. In the electric dipole approximation the interaction Hamiltonian is given by Eq. 4.56:

$$H' = -\boldsymbol{\mu}\cdot\mathbf{E} = -\mu E \tag{4.62}$$

where μ is the component of the electric dipole moment in the polarization direction of the electric field $\mathbf{E}$. Assuming that the states $|u_i\rangle$ have definite parity (we assume that the system is symmetric with respect to the centre), the matrix elements of the dipole moment operator are given by:

$$\mu = \begin{pmatrix} 0 & \mu_{12} \\ \mu_{21} & 0 \end{pmatrix} \tag{4.63}$$

where $\mu_{21} = \mu_{12}^*$. We can assume that μ_{12} and μ_{21} are real and therefore equal. The average value of the dipole moment can be calculated by using the density matrix formalism:

$$\langle\mu\rangle = \mathrm{Tr}(\rho\mu) = \mu(\rho_{12} + \rho_{21}) = 2\mu\mathrm{Re}\{\rho_{21}\}. \tag{4.64}$$

where we have used the fact that $\rho_{12} = \rho_{21}^*$. The temporal evolution of the density matrix is given by the von-Neumann equation:

$$\frac{\partial\rho}{\partial t} = \frac{i}{\hbar}[\rho, H]. \tag{4.65}$$

where the commutator is given by:

$$\begin{aligned}
[\rho, H_0 + H'] &= \begin{pmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{pmatrix}\begin{pmatrix} \mathcal{E}_1 & -\mu E \\ -\mu E & \mathcal{E}_2 \end{pmatrix} + \\
&- \begin{pmatrix} \mathcal{E}_1 & -\mu E \\ -\mu E & \mathcal{E}_2 \end{pmatrix}\begin{pmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{pmatrix} = \\
&= \begin{pmatrix} -\rho_{12}\mu E + \rho_{21}\mu E & -\rho_{12}(\mathcal{E}_1 - \mathcal{E}_2) - (\rho_{11} - \rho_{22})\mu E \\ \rho_{21}(\mathcal{E}_1 - \mathcal{E}_2) + (\rho_{11} - \rho_{22})\mu E & -\rho_{21}\mu E + \rho_{12}\mu E \end{pmatrix}
\end{aligned} \tag{4.66}$$

Therefore:

$$\begin{aligned}
\frac{d\rho_{21}}{dt} &= \frac{i}{\hbar}[-\rho_{21}(\mathcal{E}_2 - \mathcal{E}_1) + (\rho_{11} - \rho_{22})\mu E] \\
\frac{d\rho_{21}}{dt} &= -i\omega_0\rho_{21} + i\frac{\mu}{\hbar}E(t)(\rho_{11} - \rho_{22})
\end{aligned} \tag{4.67}$$

where $\hbar\omega_0 = \mathcal{E}_2 - \mathcal{E}_1$. Moreover ($\rho_{12} = \rho_{21}^*$):

$$\frac{d\rho_{11}}{dt} = i\frac{\mu}{\hbar}E(t)(\rho_{21} - \rho_{21}^*)$$
$$\frac{d\rho_{22}}{dt} = -i\frac{\mu}{\hbar}E(t)(\rho_{21} - \rho_{21}^*) \tag{4.68}$$

which give:

$$\frac{d}{dt}(\rho_{11} - \rho_{22}) = 2i\frac{\mu}{\hbar}E(t)(\rho_{21} - \rho_{21}^*). \tag{4.69}$$

4.5.1 Relaxation processes

Let us consider Eq. 4.69, which describes the temporal evolution of the population difference between the two states. Indeed, the diagonal elements of the density matrix give the probability that the system is in a given eigenstate, therefore the population of the state $|1\rangle$ is given by:

$$N_1 = N\rho_{11}, \tag{4.70}$$

where N is the number of atoms or molecules per unit volume and the population of the state $|2\rangle$ is:

$$N_2 = N\rho_{22}. \tag{4.71}$$

If the perturbation is turned off, Eq. 4.69 becomes:

$$\frac{d}{dt}(\rho_{11} - \rho_{22}) = 0, \tag{4.72}$$

so that $\rho_{11} - \rho_{22}$ remains constant. It is evident that without any external perturbation, the populations of the states have to evolve towards their thermal equilibrium values, $(\rho_{11})_{eq}$ and $(\rho_{22})_{eq}$, with a given time constant, which depends on the particular nature of the system under investigation. The population difference, $N(\rho_{11} - \rho_{22})$, has to evolve in the same way towards the thermodynamic equilibrium value, $N(\rho_{11} - \rho_{22})_{eq}$, with a time constant T_1, which is generally called *longitudinal relaxation time* or population lifetime. We can incorporate this relaxation process in Eq. 4.69 in the following way:

$$\frac{d}{dt}(\rho_{11} - \rho_{22}) = 2i\frac{\mu}{\hbar}E(t)(\rho_{21} - \rho_{21}^*) - \frac{(\rho_{11} - \rho_{22}) - (\rho_{11} - \rho_{22})_{eq}}{T_1}. \tag{4.73}$$

The longitudinal relaxation time is a measure of the time required for a system to reach the thermodynamic equilibrium with its environment. The population relaxation can be due to interactions between a molecule and its lattice, or, in the case of gases or liquids to collisions between molecules or to spontaneous emission. The temporal evolution of the off-diagonal elements of the density matrix is given by Eq. 4.67. If the perturbation is turned off this equation becomes:

$$\frac{d\rho_{21}}{dt} = -i\omega_0\rho_{21} \tag{4.74}$$

so that $\rho_{21}(t) = \rho_{21}(0)\exp(-i\omega_0 t)$, and the off-diagonal elements continue to oscillate with angular frequency $\omega_0 = (\mathcal{E}_2 - \mathcal{E}_1)/\hbar$. Equation 4.67 does not contain

any term leading to dephasing of the dipoles generated by the perturbation. At thermal equilibrium the off-diagonal elements of ρ must be zero as a result of several dephasing processes. Indeed, any process which contributes to the relaxation of the population also contributes to dephasing. Moreover, there are processes which contribute to dephasing but do not alter the population difference: for example an elastic collision, in which one molecule is excited from the state $|u_1\rangle$ to the state $|u_2\rangle$ and another makes the inverse transition, from $|u_2\rangle$ to $|u_1\rangle$, does not change the population difference but can lead to a dephasing of the off-diagonal elements. Therefore, if all the molecules are initially vibrating in phase, after a given time they will vibrate with arbitrary phases relative to each other. The dephasing process can be added phenomenologically in Eq. 4.67 as follows:

$$\frac{d\rho_{21}}{dt} = -i\omega_0\rho_{21} + i\frac{\mu}{\hbar}(\rho_{11} - \rho_{22})E(t) - \frac{\rho_{21}}{T_2} \tag{4.75}$$

where T_2 is called *transverse relaxation time*. We will see that T_2 is associated with the linewidth of a transition. Since all processes contributing to T_1 also contribute to T_2 and there are dephasing processes, which contribute only to T_2, in general $T_2 < T_1$. We are now ready to solve Eqs. 4.73 and 4.75 by assuming a harmonic perturbation.

4.5.2 Harmonic perturbation

Let us consider a monochromatic electromagnetic field with electric field:

$$E(t) = E_0 \cos\omega t = \frac{E_0}{2}(e^{i\omega t} + e^{-i\omega t}) \tag{4.76}$$

We will assume that $\omega \approx \omega_0$. Since in the field-free case the element ρ_{21} oscillates with frequency ω_0, as shown by Eq. 4.74, it is useful to transform into the rotating frame ($\omega \approx \omega_0$):

$$\rho_{21}(t) = \sigma_{21}(t)\, e^{-i\omega t} \tag{4.77}$$

$$\rho_{12}(t) = \rho_{21}^*(t) = \sigma_{21}^*\, e^{i\omega t} \tag{4.78}$$

where $\sigma_{21}(t)$ is a slowly varying function. Equation 4.75 can be rewritten as:

$$\begin{aligned}\frac{d\sigma_{21}}{dt}\, e^{-i\omega t} - i\omega\sigma_{21}e^{-i\omega t} = \\ = -i\omega_0\sigma_{21}e^{-i\omega t} + \frac{\mu}{\hbar}(\rho_{11} - \rho_{22})\frac{E_0}{2}(e^{i\omega t} + e^{-i\omega t}) - \frac{\sigma_{21}}{T_2}e^{-i\omega t}\end{aligned} \tag{4.79}$$

if we consider only the terms with $e^{-i\omega t}$ we have:

$$\frac{d\sigma_{21}}{dt} = i(\omega - \omega_0)\sigma_{21} + i\frac{\mu E_0}{2\hbar}(\rho_{11} - \rho_{22}) - \frac{\sigma_{21}}{T_2} \tag{4.80}$$

Equation 4.73 can be written as:

$$\frac{d}{dt}(\rho_{11} - \rho_{22}) = \frac{i\mu E_0}{\hbar}(\sigma_{21} - \sigma_{21}^*) - \frac{(\rho_{11} - \rho_{22}) - (\rho_{11} - \rho_{22})_{eq}}{T_1} \tag{4.81}$$

where we have kept only the terms with no exponential time dependence. We have neglected the terms with $e^{2i\omega t}$ and $e^{-2i\omega t}$. From the physical point of view this is reasonable since the contribution of these terms averages to zero on a temporal interval shorter than that of interest.

The average value of the dipole moment (Eq. 4.64) is given by:

$$\begin{aligned}\langle\mu\rangle &= 2\mu\mathrm{Re}\{\rho_{21}\} = 2\mu\mathrm{Re}\{\sigma_{21}e^{-i\omega t}\} = \\ &= 2\mu[\mathrm{Re}\sigma_{21}(t)\cos\omega t + \mathrm{Im}\sigma_{21}(t)\sin\omega t]\end{aligned} \tag{4.82}$$

We can now calculate the steady-state solution of Eqs. 4.80 and 4.81.

Steady-state solutions

To find the steady-state solutions of Eqs. 4.80 and 4.81 we put:

$$\frac{d\sigma_{21}}{dt} = \frac{d}{dt}(\rho_{11} - \rho_{22}) = 0. \tag{4.83}$$

Therefore:

$$i(\omega - \omega_0)\sigma_{21} + \frac{i\mu E_0}{2\hbar}(\rho_{11} - \rho_{22}) - \frac{\sigma_{21}}{T_2} = 0 \tag{4.84}$$

$$\frac{i\mu E_0}{\hbar}(\sigma_{21} - \sigma_{21}^*) - \frac{(\rho_{11} - \rho_{22}) - (\rho_{11} - \rho_{22})_{eq}}{T_1} = 0 \tag{4.85}$$

If we take the complex conjugate of Eq. 4.84:

$$-i(\omega - \omega_0)\sigma_{21}^* - \frac{i\mu E_0}{2\hbar}(\rho_{11} - \rho_{22}) - \frac{\sigma_{21}^*}{T_2} = 0 \tag{4.86}$$

and we sum Eqs. 4.84 and 4.86 we obtain:

$$i(\omega - \omega_0)(\sigma_{21} - \sigma_{21}^*) - \frac{\sigma_{21} + \sigma_{21}^*}{T_2} = 0 \tag{4.87}$$

If we subtract Eqs. 4.84 and 4.86 we obtain:

$$i(\omega - \omega_0)(\sigma_{21} + \sigma_{21}^*) + \frac{i\mu E_0}{\hbar}(\rho_{11} - \rho_{22}) - \frac{\sigma_{21} - \sigma_{21}^*}{T_2} = 0 \tag{4.88}$$

These last two equations can be written as:

$$i(\omega - \omega_0)(2i\ \mathrm{Im}\sigma_{21}) - \frac{2\ \mathrm{Re}\sigma_{21}}{T_2} = 0 \tag{4.89}$$

$$i(\omega - \omega_0)2\ \mathrm{Re}\sigma_{21} + \frac{i\mu E_0}{\hbar}(\rho_{11} - \rho_{22}) - \frac{2i\ \mathrm{Im}\sigma_{21}}{T_2} = 0 \tag{4.90}$$

With suitable calculations:

$$\mathrm{Re}\sigma_{21} = -(\omega - \omega_0)T_2\ \mathrm{Im}\sigma_{21} \tag{4.91}$$

$$i(\omega - \omega_0)[-2(\omega - \omega_0)T_2\ \mathrm{Im}\sigma_{21}] + \frac{i\mu E_0}{\hbar}(\rho_{11} - \rho_{22}) - \frac{2i}{T_2}\ \mathrm{Im}\sigma_{21} = 0 \tag{4.92}$$

from which it is possible to obtain:

$$\mathrm{Im}\sigma_{21} = \frac{\mu E_0(\rho_{11} - \rho_{22})/\hbar}{2[(\omega - \omega_0)^2T_2^2 + 1]/T_2} = \frac{\Omega T_2(\rho_{11} - \rho_{22})}{1 + (\omega - \omega_0)^2T_2^2} \tag{4.93}$$

where $\Omega = \mu E_0/2\hbar$. We can now rewrite Eq. 4.85 as:

$$\frac{i\mu E_0}{\hbar}2i\ \mathrm{Im}\sigma_{21} - \frac{(\rho_{11} - \rho_{22}) - (\rho_{11} - \rho_{22})_{eq}}{T_1} = 0 \tag{4.94}$$

from which we obtain:

$$\rho_{11} - \rho_{22} = -2\frac{\mu E_0}{\hbar}\,\mathrm{Im}\sigma_{21}\,T_1 + (\rho_{11} - \rho_{22})_{eq} \tag{4.95}$$

which can be used in Eq. 4.93, so that :

$$\mathrm{Im}\sigma_{21} = \frac{\Omega\,T_2(\rho_{11} - \rho_{22})_{eq}}{1 + (\omega - \omega_0)^2T_2^2 + 4\Omega^2T_2T_1} \tag{4.96}$$

$$\mathrm{Re}\sigma_{21} = \frac{(\omega_0 - \omega)T_2^2\Omega(\rho_{11} - \rho_{22})_{eq}}{1 + (\omega - \omega_0)^2T_2^2 + 4\Omega^2T_2T_1} \tag{4.97}$$

$$(\rho_{11} - \rho_{22}) = (\rho_{11} - \rho_{22})_{eq}\frac{1 + (\omega - \omega_0)^2T_2^2}{1 + (\omega - \omega_0)^2T_2^2 + 4\Omega^2T_2T_1} \tag{4.98}$$

The macroscopic polarization can be written as follows:

$$P = N\langle\mu\rangle = \frac{\mu^2\Delta N_0T_2}{\hbar}\,E_0\frac{\sin\omega t + (\omega_0 - \omega)T_2\,\cos\omega t}{1 + (\omega - \omega_0)^2T_2^2 + 4\Omega^2T_2T_1} \tag{4.99}$$

and the population difference per unit volume is:

$$\Delta N = \Delta N_0\frac{1 + (\omega - \omega_0)^2T_2^2}{1 + (\omega - \omega_0)^2T_2^2 + 4\Omega^2T_2T_1} \tag{4.100}$$

where $\Delta N_0 = N(\rho_{11} - \rho_{22})_{eq}$ is the unperturbed population difference. The polarization P can be written as:

$$P = \frac{\pi\mu^2}{\hbar}\Delta N\;E_0\;g(\omega - \omega_0)[\sin\omega t + (\omega_0 - \omega)T_2\;\cos\omega t] \tag{4.101}$$

where:

$$g(\omega - \omega_0) = \frac{T_2}{\pi[1 + (\omega - \omega_0)^2T_2^2]} \tag{4.102}$$

is the normalized Lorentzian lineshape function. If we now write the atomic susceptibility χ as:

$$\chi = \chi' + i\chi'' \tag{4.103}$$

then:

$$P(t) = \mathrm{Re}\{\epsilon_0\chi E_0e^{-i\omega t}\} = E_0(\epsilon_0\chi'\cos\omega t + \epsilon_0\chi''\sin\omega t) \tag{4.104}$$

From Eqs. 4.101 and 4.104 we obtain:

$$\chi'(\omega) = \frac{\pi\mu^2}{\epsilon_0\hbar}(\omega_0 - \omega)T_2\Delta Ng(\omega - \omega_0) \tag{4.105}$$

$$\chi''(\omega) = \frac{\pi\mu^2}{\epsilon_0\hbar}\Delta Ng(\omega - \omega_0) \tag{4.106}$$

Lorentzian function

A few remarks about the lineshape function $g(\omega - \omega_0)$. If we want to write the lineshape function as a function of frequency ν we have to recall that $g(\omega)d\omega = g(\nu)d\nu$, so that:

$$g(\nu - \nu_0) = \frac{2T_2}{1 + 4\pi^2(\nu - \nu_0)^2 T_2^2} \tag{4.107}$$

The full-width at half-maximum of the functions $g(\omega)$ and $g(\nu)$ are:

$$\Delta\omega = \frac{2}{T_2} \qquad \Delta\nu = \frac{1}{\pi T_2} \tag{4.108}$$

thus showing that the transverse relaxation time is related to the linewidth of the transition. Moreover the lineshape function is normalized:

$$\int_{-\infty}^{\infty} g(\nu)\, d\nu = 1 \tag{4.109}$$

Figure 4.1 shows the real and imaginary components of the atomic susceptibility as a function of the frequency detuning $(\omega - \omega_0)$ in the limit $4\Omega^2 T_2 T_1 \ll 1$.

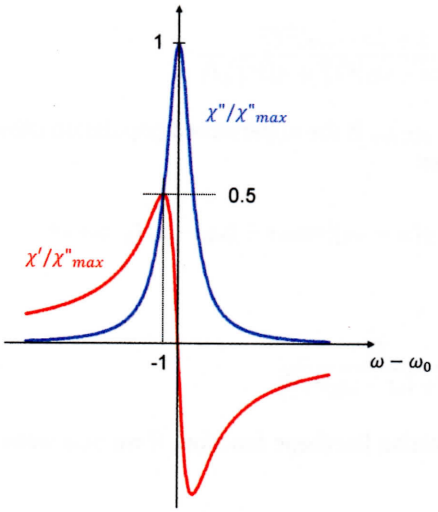

Figure 4.1: Real (red line) and imaginary (blue line) components of the atomic susceptibility as a function of the frequency detuning $\omega - \omega_0$. Both functions are normalized to the maximum value of the imaginary component.

4.6 From optical susceptibility to absorption coefficient

It is well known that the absorption coefficient of a medium can be obtained from the atomic susceptibility. The macroscopic polarization $\mathbf{P} = N\langle\boldsymbol{\mu}\rangle$ can be decomposed in two components:

$$\mathbf{P}_T = \mathbf{P}_{nr} + \mathbf{P}_r \tag{4.110}$$

where $\mathbf{P}_r$ is the polarization vector of the system near resonance and $\mathbf{P}_{nr}$ is the non-resonant contribution, which corresponds to all the off-resonant contributions:

$$\mathbf{P}_{nr} = \epsilon_0 \chi_{nr} \mathbf{E}, \tag{4.111}$$

where χ_{nr} is the non-resonant component of the atomic susceptibility.

$$\mathbf{P}_r = \epsilon_0 \chi_r(\omega)\mathbf{E} = \epsilon_0 \chi(\omega)\mathbf{E} \tag{4.112}$$

where $\chi(\omega)$ is the resonant component of the atomic susceptibility, which we have calculated in the previous section. While χ_{nr} is weakly dependent on frequency, $\chi(\omega)$ strongly depends on ω, as shown in Fig. 4.1. Therefore:

$$\mathbf{P}_T = \epsilon_0[\chi_{nr} + \chi(\omega)]\mathbf{E} \tag{4.113}$$

The displacement vector $\mathbf{D}$ is given by:

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P}_T = \epsilon_0[1 + \chi_{nr} + \chi(\omega)]\mathbf{E} \tag{4.114}$$

We can define the total dielectric constant, ϵ_{tot}, as:

$$\epsilon_{tot} = \epsilon_0[1 + \chi_{nr} + \chi(\omega)] = \epsilon_0 \epsilon_{r,tot}. \tag{4.115}$$

The total refractive index is given by:

$$n_{tot}^2 = \epsilon_{r,tot} = (1 + \chi_{nr}) + \chi(\omega) = n^2 + \chi(\omega) = n^2 \left[1 + \frac{\chi(\omega)}{n^2}\right] \tag{4.116}$$

where n is the refractive index due to all the off-resonant transitions (i.e., it is the refractive index far from the resonance). Therefore:

$$k_{tot}^2 = \left(\frac{\omega n_{tot}}{c}\right)^2 = \left(\frac{\omega n}{c}\right)^2 \left[1 + \frac{\chi(\omega)}{n^2}\right] \tag{4.117}$$

If $\chi(\omega)/n^2 \ll 1$:

$$k_{tot} \simeq \frac{\omega n}{c}\left[1 + \frac{\chi(\omega)}{2n^2}\right], \qquad n_{tot} \simeq n_0\left[1 + \frac{\chi(\omega)}{2n^2}\right] \tag{4.118}$$

$$k_{tot} = k' + ik'' \simeq k\left[1 + \frac{\chi'(\omega)}{2n^2}\right] + i\frac{k\chi''(\omega)}{2n^2} \tag{4.119}$$

where $k = \omega n/c$. The electric field of the electromagnetic field, which is propagating in the medium is:

$$\begin{aligned} E(z,t) &= \mathrm{Re}\{E_0 e^{i(k_{tot}z - \omega t)}\} \\ &= \mathrm{Re}\left\{E_0 \exp i\left[k\left(1 + \frac{\chi'(\omega)}{2n^2}\right)z - \omega t\right] \exp\left(-\frac{k\chi''(\omega)}{2n^2}z\right)\right\} \end{aligned} \tag{4.120}$$

Therefore the polarization of the medium gives rise to a variation of the phase delay per unit length from k to $k + \Delta k$, where:

$$\Delta k = \frac{k\chi'(\omega)}{2n^2} \tag{4.121}$$

and an exponential variation of the amplitude, $\exp(-\alpha z/2)$, where:

$$\alpha(\omega) = \frac{k\chi''(\omega)}{n^2} \tag{4.122}$$

By using the expression (Eq. 4.106) of $\chi''(\omega)$ obtained by using the density matrix formalism in Eq. 4.122, the absorption coefficient can be written as:

$$\alpha(\omega) = \frac{k\mu_{12}^2\pi}{\epsilon_0 n^2 \hbar}\Delta N g(\omega - \omega_0), \tag{4.123}$$

we recall that $\Delta N = N_1 - N_2$. If $N_2 > N_1$ the system is characterized by a negative absorption, i.e., it presents an optical gain. In the following chapter we will apply the main results of this chapter to semiconductors.

5. Optical properties of semiconductors

5.1 Stimulated transitions: selection rules

In Chapter 3 we have obtained the Fermi's Golden rule, which, in the case of a two-level system ($m = 1$, $k = 2$), can be written as:

$$W_{12} = \frac{2\pi}{\hbar^2}|H'_{21}|^2\delta(\omega - \omega_0) \tag{5.1}$$

where $\hbar\omega_0 = \mathcal{E}_2 - \mathcal{E}_1$. Within the electric dipole approximation, the interaction Hamiltonian can be written as:

$$H' = -\boldsymbol{\mu} \cdot \mathbf{E} \tag{5.2}$$

and the matrix element H'_{21} can be calculated as follows:

$$H'_{21} = \langle\psi_2| - \boldsymbol{\mu} \cdot \mathbf{E}|\psi_1\rangle \tag{5.3}$$

where $|\psi_1\rangle$ and $|\psi_2\rangle$ are the Bloch wavefunctions of the states with energy $\mathcal{E}_1$ and $\mathcal{E}_2$, respectively:

$$|\psi_{1,2}(\mathbf{r})\rangle = |u_{1,2}(\mathbf{r})\rangle e^{i\mathbf{k}_{1,2}\cdot\mathbf{r}} \tag{5.4}$$

$\mathbf{E}(\mathbf{r}, t)$ is the electric field, which will be assumed polarized along the direction indicated by the unit vector $\hat{e}$:

$$\mathbf{E}(\mathbf{r}, t) = \frac{E(\mathbf{r}_0)}{2}\,\hat{e}\,e^{i(\mathbf{k}_{opt}\cdot\mathbf{r}-\omega t)} \tag{5.5}$$

where $\mathbf{k}_{opt}$ is the field wavevector. The matrix element H'_{21} is given by:

$$H'_{21} = \langle u_2 e^{-i\mathbf{k}_2\cdot\mathbf{r}}| - \boldsymbol{\mu} \cdot \frac{E(\mathbf{r}_0)}{2}\,\hat{e}\,e^{i(\mathbf{k}_{opt}\cdot\mathbf{r}-\omega t)}|u_1 e^{i\mathbf{k}_1\cdot\mathbf{r}}\rangle \tag{5.6}$$

Since the term $\exp[i(\mathbf{k}_1 + \mathbf{k}_{opt} - \mathbf{k}_2) \cdot \mathbf{r}]$ oscillates rapidly with $\mathbf{r}$, H'_{21} is zero unless

$$\mathbf{k}_2 = \mathbf{k}_{opt} + \mathbf{k}_1 \tag{5.7}$$

Since $\hbar\mathbf{k}_{1,2}$ are the electron momenta in the valence and conduction band and $\hbar\mathbf{k}_{opt}$ is the photon momentum, Eq. 5.7 shows that total momentum is conserved in the transition. If we assume that the refractive index of the semiconductor is $n = 3.5$ and $\lambda = 1$ μm, we have $k_{opt} = 2\pi n/\lambda \simeq 2 \times 10^5$ cm^{-1}. On the other hand, typical values for $k_{1,2} = k_{v,c}$ are of the order of $10^7 - 10^8$ cm^{-1} (see Exercise 5.1). Therefore $k_{opt} \ll k_{c,v}$ and that the momentum conservation 5.7 can be written as:

$$\mathbf{k}_1 = \mathbf{k}_2 \tag{5.8}$$

This equation is often referred to as the **k**-selection or **k**-conservation rule, and it indicates that stimulated transitions must occur vertically in the **k**−space. Note that the **k**-selection rule can be derived also considering the interaction Hamiltonian $H' = (e/m)\mathbf{A} \cdot \mathbf{p}$.
Since the transition rate depends on $\delta(\omega - \omega_0)$, we can conclude that $\omega = \omega_0$. This means

$$\hbar\omega = \mathcal{E}_2 - \mathcal{E}_1 \tag{5.9}$$

which represents the energy conservation rule. Finally, we note that the electromagnetic wave does not interact with the electron spin, so that the spin is not involved in the interaction Hamiltonian. We can conclude that the spin cannot change in the transition and the spin-selection rule can be written as:

$$\Delta S = 0 \tag{5.10}$$

Exercise 5.1 Calculate typical values of **k** for a thermal electron in bulk GaAs ($m_c = 0.067\ m_0$, $m_v = m_{hh} = 0.45\ m_0$).

For an electron in the conduction band having thermal velocity v_{th} we have (see Sect. 2.7):

$$\frac{1}{2} m_c v_{th}^2 = \frac{3}{2} k_B T$$

where T is the electron temperature. We also have

$$p = \hbar k_c = m_c v_{th}$$

Combining the two previous expressions, we obtain

$$k_c = (3 m_c k_B T)^{1/2}/\hbar$$

In the case of bulk GaAs at $T = 300$ K ($k_B T = 26$ meV) we obtain $k_c = 2.6 \times 10^6$ cm^{-1} and $k_v = (m_v/m_c)^{1/2} k_c = 6.7 \times 10^6$ cm^{-1}. ■

5.2 Joint density of states

Let us consider two energy levels, $\mathcal{E}_1$ (in the valence band) and $\mathcal{E}_2$ (in the conduction band), with energy difference $\mathcal{E}_2 - \mathcal{E}_1 = \mathcal{E}_0 = \hbar\omega_0$, where ω_0 is the angular frequency

of the transition. By using the parabolic approximation for the dispersion curve we can write:

$$\mathcal{E}_1 = -\frac{\hbar^2 k^2}{2m_v} \tag{5.11}$$

$$\mathcal{E}_2 = \mathcal{E}_g + \frac{\hbar^2 k^2}{2m_c} \tag{5.12}$$

where $\mathcal{E}_g$ is the energy gap. In Eqs. 5.11 and 5.12 we have assumed $k_c = k_v = k$. The energy difference is:

$$\mathcal{E}_0 = \mathcal{E}_2 - \mathcal{E}_1 = \mathcal{E}_g + \frac{\hbar^2 k^2}{2}\left(\frac{1}{m_c} + \frac{1}{m_v}\right) = \mathcal{E}_g + \frac{\hbar^2 k^2}{2m_r} \tag{5.13}$$

where m_r is the *reduced mass*:

$$\frac{1}{m_r} = \frac{1}{m_c} + \frac{1}{m_v} \tag{5.14}$$

The energies $\mathcal{E}_1$ and $\mathcal{E}_2$ are related to the transition energy $\mathcal{E}_0$ by the following expressions:

$$\mathcal{E}_1 = -\frac{m_r}{m_v}(\mathcal{E}_0 - \mathcal{E}_g) = -\frac{m_r}{m_v}(\hbar\omega_0 - \mathcal{E}_g) \tag{5.15}$$

$$\mathcal{E}_2 = \mathcal{E}_g + \frac{m_r}{m_c}(\mathcal{E}_0 - \mathcal{E}_g) = \mathcal{E}_g + \frac{m_r}{m_c}(\hbar\omega_0 - \mathcal{E}_g) \tag{5.16}$$

The *joint density of states* $\rho_j(\mathcal{E}_0)$ is defined as follows: $\rho_j(\mathcal{E}_0)d\mathcal{E}_0$ is the number of transitions per unit volume with transition energy in the range between $\mathcal{E}_0$ and $\mathcal{E}_0 + d\mathcal{E}_0$. As a consequence of the selection rules a given state in the valence band is associated to a single state in the conduction band, with the same k value and the same spin. Therefore the number of transitions is equal to the number of the corresponding states in the valence or in the conduction bands:

$$\rho_j(\mathcal{E}_0)d\mathcal{E}_0 = \rho(k)dk = \frac{k^2}{\pi^2}\,dk \tag{5.17}$$

where we have used Eq. 2.22. From Eq. 5.13 we have:

$$d\mathcal{E}_0 = \frac{\hbar^2 k}{m_r}dk \tag{5.18}$$

so that:

$$\rho_j(\mathcal{E}_0) = \frac{m_r k}{\pi^2\hbar^2} \tag{5.19}$$

By using the expression of k from Eq. 5.13 in Eq. 5.19 we have:

$$\rho_j(\mathcal{E}_0) = \frac{1}{2\pi^2}\left(\frac{2m_r}{\hbar^2}\right)^{3/2}(\mathcal{E}_0 - \mathcal{E}_g)^{1/2} \tag{5.20}$$

The joint density of states as a function of the transition frequency can be calculated by writing: $\rho_j(\omega_0)d\omega_0 = \rho_j(\mathcal{E}_0)d\mathcal{E}_0$, where $d\mathcal{E}_0 = \hbar d\omega_0$, so that:

$$\rho_j(\omega_0) = \frac{1}{2\pi^2\hbar^2}(2m_r)^{3/2}(\hbar\omega_0 - \mathcal{E}_g)^{1/2} \tag{5.21}$$

In order to have an absorption transition the state with energy $\mathcal{E}_1$ must be occupied by an electron, while the state with energy $\mathcal{E}_2$ must be empty. Therefore the number of absorption transitions per unit volume is given by:

$$dN_{12} = \rho_j(\omega_0) f_v(\mathcal{E}_1)[1 - f_c(\mathcal{E}_2)]d\omega_0 \tag{5.22}$$

We assume that thermal equilibrium has been achieved (separately) in the conduction and valence bands so that quasi-Fermi levels are used in Eq. 5.22. Stimulated emission events have to be taken into account from the state $\mathcal{E}_2$ to the state $\mathcal{E}_1$. The number of stimulated emission transitions per unit volume is given by:

$$dN_{21} = \rho_j(\omega_0) f_c(\mathcal{E}_2)[1 - f_v(\mathcal{E}_1)]d\omega_0 \tag{5.23}$$

Therefore, the net number of absorption transitions per unit volume is:

$$dN = dN_{12} - dN_{21} = \rho_j(\omega_0)[f_v(\mathcal{E}_1) - f_c(\mathcal{E}_2)]d\omega_0 \tag{5.24}$$

Since $\mathcal{E}_1$ and $\mathcal{E}_2$ are determined by $\mathcal{E}_0$ (see Eqs. 5.15 and 5.16), the previous equation can be written as:

$$dN = \rho_j(\omega_0)[f_v(\hbar\omega_0) - f_c(\hbar\omega_0)]d\omega_0 \tag{5.25}$$

5.3 Susceptibility and absorption coefficient in a semiconductor

In the previous Chapter we have seen that the absorption coefficient of a medium is related to the imaginary part of the optical susceptibility by the following expression:

$$\alpha(\omega) = \frac{k\chi''(\omega)}{n^2} = \frac{\omega\chi''(\omega)}{nc} \tag{5.26}$$

where $\chi''(\omega)$ is given by (see Eq. 4.106)

$$\chi''(\omega) = \frac{\pi\mu_{21}^2}{\epsilon_0\hbar}\Delta N g(\omega - \omega_0) \tag{5.27}$$

In the case of a semiconductor, ΔN must be replaced by the net number of absorption transitions per unit volume, dN, given by Eq. 5.25, so that the absorption coefficient due to the transitions with transition frequency in the range between ω_0 and $\omega_0 + d\omega_0$ is:

$$d\alpha = \frac{\omega\pi}{nc\epsilon_0\hbar}\mu_{21}^2\rho_j(\omega_0)g(\omega - \omega_0)[f_v(\hbar\omega_0) - f_c(\hbar\omega_0)]d\omega_0 \tag{5.28}$$

The absorption coefficient of the semiconductor can be obtained from Eq. 5.28 by integrating over all possible transition frequencies ω_0:

$$\alpha(\omega) = \frac{\omega\pi\mu_{21}^2}{nc\epsilon_0\hbar}\int_{\mathcal{E}_g/\hbar}^{\infty}\rho_j(\omega_0)\,[f_v(\hbar\omega_0) - f_c(\hbar\omega_0)]g(\omega - \omega_0)\,d\omega_0 \tag{5.29}$$

In Eq. 5.29 the matrix element μ_{21} has been assumed independent of ω_0. This integral can be easily calculated if we assume that the Lorentzian function $g(\omega - \omega_0)$ behaves as a delta-function $\delta(\omega - \omega_0)$ in comparison to the slowly varying functions $\rho_j(\omega_0)$ and $f_v(\hbar\omega_0) - f_c(\hbar\omega_0)$, so that:

$$\alpha(\omega) = \frac{\omega\pi}{cn\epsilon_0\hbar}\mu_{21}^2\rho_j(\omega)\,[f_v(\hbar\omega) - f_c(\hbar\omega)] = \alpha_0(\omega)[f_v(\hbar\omega) - f_c(\hbar\omega)]. \tag{5.30}$$

In a semiconductor at thermal equilibrium at $T = 0$ K, quasi-Fermi levels coincide with the Fermi level, $f_v(\hbar\omega) = 1$ and $f_c(\hbar\omega) = 0$, so that $\alpha(\omega) = \alpha_0(\omega)$, which represents the maximum absorption coefficient at a given frequency. Note that in the case of an intrinsic semiconductor, assuming $\mathcal{E}_g \gg k_B T$, $f_v(\hbar\omega) \simeq 1$ and $f_c(\hbar\omega) \simeq 0$, so that $\alpha(\omega) \simeq \alpha_0(\omega)$ even at room temperature.

The matrix element μ_{21} can be calculated as follows. We have first to write the relation between the matrix elements $\mathbf{r}_{cv}$ and $\mathbf{p}_{cv}$, given by the following expressions:

$$\mathbf{r}_{cv} = \langle u_{c,\mathbf{k}} | \mathbf{r} | u_{v,\mathbf{k}} \rangle \tag{5.31}$$

$$\mathbf{p}_{cv} = \langle u_{c,\mathbf{k}} | \mathbf{p} | u_{v,\mathbf{k}} \rangle \tag{5.32}$$

These two matrix elements are related to the interaction Hamiltonians $-\boldsymbol{\mu} \cdot \mathbf{E}$ and $-(q/m)\mathbf{A} \cdot \mathbf{p}$, respectively. It is possible to demonstrate (see the following gray box) that:

$$\mathbf{r}_{cv} = -\frac{i\hbar}{m_0(\mathcal{E}_c - \mathcal{E}_v)} \mathbf{p}_{cv} \tag{5.33}$$

where m_0 is the free electron mass. We have already mentioned in Ch. 1 that the Kane energy $\mathcal{E}_P$ is given by (see Eq. 1.107)

$$\mathcal{E}_P = \frac{2}{m_0} |\mathbf{p}_{cv}|^2, \tag{5.34}$$

which is largely constant in the case of III-V semiconductors (values for the Kane energy for various semiconductors are given in Table 5.1).

	GaAs	**InP**	**InAs**
$\mathcal{E}_g$ (eV)	1.52	1.42	0.43
$\mathcal{E}_P$ (eV)	22.71	17	21.11
r_{cv} (Å)	6.12	5.67	20.9

Table 5.1: Kane energy ($\mathcal{E}_P$) and matrix element r_{cv} for various III-V semiconductors at $T = 0$ K (adapted from G. Bastard, Wave Mechanics applied to Semiconductor Heterostructures, Wiley, New York (1991)).

To first order in k, the matrix elements $\mathbf{p}_{cv}$ are constant. By using Eqs. 5.33 and 5.34 we can write:

$$|\mathbf{r}_{cv}| = \frac{\hbar}{\mathcal{E}_g} \sqrt{\frac{\mathcal{E}_P}{2m_0}}. \tag{5.35}$$

In order to take into account the fact that in semiconductors only two valence bands out of three effectively contribute to optical transitions, namely the heavy- and light-hole bands (since the split-off energy Δ_{SO} is generally much larger than $k_B T$) it is common practice to use the matrix element x_{cv}, defined in the following way:

$$x_{cv}^2 = \frac{2}{3} |\mathbf{r}_{cv}|^2 = \frac{\hbar^2 \mathcal{E}_P}{3m_0 \mathcal{E}_g^2}. \tag{5.36}$$

The matrix element of the electric dipole moment can be written as:

$$\mu_{21}^2 = e^2 x_{cv}^2. \tag{5.37}$$

The absorption coefficient given by Eq. 5.30 can be written as follows:

$$\alpha(\omega) = \frac{\omega\pi e^2}{cn\epsilon_0\hbar} x_{cv}^2 \rho_j(\omega)[f_v(\hbar\omega) - f_c(\hbar\omega)]. \tag{5.38}$$

The absorption coefficient can be written as a function of the photon energy, $\hbar\omega$:

$$\alpha(\hbar\omega) = \frac{\omega\pi e^2}{cn\epsilon_0} x_{cv}^2 \rho_j(\hbar\omega)[f_v(\hbar\omega) - f_c(\hbar\omega)]. \tag{5.39}$$

5.3.1 Relation between $\mathbf{r}_{cv}$ and $\mathbf{p}_{cv}$.

We will assume that the electric field $\mathbf{E}$ and the vector potential $\mathbf{A}$ are polarized along the $z-$axis.

$$\mathbf{A}(t) = \frac{A(\mathbf{r}_0)}{2}\mathbf{u}_z\, e^{i(\mathbf{k}_{opt}\cdot\mathbf{r}-\omega t)} \tag{5.40}$$

The matrix element, H'_{21}, of the interaction Hamiltonian $H' = (e/m)\mathbf{A}\cdot\mathbf{p}$ is given by:

$$H'_{21} = \langle\psi_2|\frac{e}{2m_0}A(\mathbf{r}_0)e^{i\mathbf{k}_{opt}\cdot\mathbf{r}}\mathbf{u}_z\cdot\mathbf{p}|\psi_1\rangle \tag{5.41}$$

By taking into account the $k-$selection rule ($\mathbf{k}_2 = \mathbf{k}_1 + \mathbf{k}_{opt}$), Eq. 5.41 can be re-written as:

$$H'_{21} = \frac{e}{2m_0}A(\mathbf{r}_0)\langle u_2|p_z|u_1\rangle. \tag{5.42}$$

where $p_z = \mathbf{p}\cdot\mathbf{u}_z$ is the component of $\mathbf{p}$ on the polarization direction.
In the same way the matrix element of the interaction Hamiltonian $H' = -\boldsymbol{\mu}\cdot\mathbf{E}$ is given by:

$$H'_{21} = \frac{e}{2}E(\mathbf{r}_0)\langle u_2|z|u_1\rangle, \tag{5.43}$$

so that

$$|H'_{21}|^2 = \frac{E^2(\mathbf{r}_0)}{4}|\mu_{21}|^2. \tag{5.44}$$

The relation between $\langle u_2|z|u_1\rangle$ and $\langle u_2|p_z|u_1\rangle$ can be calculated by using the following procedure. We first calculate the commutator $[z, H_0]$:

$$\begin{aligned}[z, H_0] &= \left[z, \frac{p_z^2}{2m_0}\right] = \frac{1}{2m_0}(zp_z^2 - p_z^2 z) = \frac{1}{2m_0}(zp_z^2 - p_z z p_z + p_z z p_z - p_z^2 z) = \\ &= \frac{1}{2m_0}([z, p_z]p_z + p_z[z, p_z])\end{aligned} \tag{5.45}$$

where the commutator $[z, p_z]$ is simply given by:

$$[z, p_z] = i\hbar \tag{5.46}$$

so that, from Eq. 5.45:

$$[z, H_0] = i\hbar \frac{p_z}{m_0}, \qquad p_z = \frac{m_0}{i\hbar}[z, H_0]. \tag{5.47}$$

We can now calculate the matrix element $\langle u_2|p_z|u_1\rangle$:

$$\begin{aligned}\langle u_2|p_z|u_1\rangle &= \frac{m_0}{i\hbar}\langle u_2|[z, H_0]|u_1\rangle = \frac{m_0}{i\hbar}\left[\langle u_2|zH_0|u_1\rangle - \langle u_2|H_0 z|u_1\rangle\right] = \\ &= \frac{m_0}{i\hbar}(\mathcal{E}_1 - \mathcal{E}_2)\langle u_2|z|u_1\rangle \end{aligned} \tag{5.48}$$

from this last equation we obtain Eq. 5.33:

$$\langle u_2|z|u_1\rangle = -\frac{i\hbar}{m_0(\mathcal{E}_2 - \mathcal{E}_1)}\langle u_2|p_z|u_1\rangle \tag{5.49}$$

From this result we can also determine the relation between the two expressions of the interaction Hamiltonian. Indeed:

$$\begin{aligned}\frac{e}{2m_0}A(\mathbf{r}_0)\langle u_2|p_z|u_1\rangle &= i\frac{e}{2m_0}\frac{E(\mathbf{r}_0)}{\omega}\frac{m_0(\mathcal{E}_2 - \mathcal{E}_1)}{\hbar}\langle u_2|z|u_1\rangle = \\ &= i\frac{\omega_0}{\omega}\left[\frac{e}{2}E(\mathbf{r}_0)\langle u_2|z|u_1\rangle\right] \end{aligned} \tag{5.50}$$

where $\hbar\omega_0 = \mathcal{E}_2 - \mathcal{E}_1$. Therefore the two expressions of the interaction Hamiltonian are equivalent only at resonance. If $\omega = \omega_0$ we obtain:

$$|p_{21}|^2 = \frac{m_0^2\omega^2|\mu_{21}|^2}{e^2}. \tag{5.51}$$

■

Exercise 5.2 Plot the absorption coefficient of bulk GaAs at room temperature assuming the following data: $\mathcal{E}_g = 1.424$ eV, $m_c = 0.067\ m_0$, $m_v = 0.45\ m_0$, $n = 3.64$. In order to fit the experimental data it is worth to assume $x_{cv} = 3.2$ Å.

From Eq. 5.38 we have:

$$\alpha_0(\omega) = \frac{\omega e^{5/2}x_{cv}^2}{2\pi c n \epsilon_0 \hbar^3}(2m_r)^{3/2}\sqrt{\frac{\hbar\omega - \mathcal{E}_g}{e}} = \mathcal{A}\sqrt{\frac{\hbar\omega - \mathcal{E}_g}{e}}$$

where

$$\mathcal{A} = \frac{\omega e^{5/2}x_{cv}^2}{2\pi c n \epsilon_0 \hbar^3}(2m_r)^{3/2} = \frac{e^{5/2}x_{cv}^2}{\lambda_0 n \epsilon_0 \hbar^3}(2m_r)^{3/2}$$

The reduced mass is $m_r = 0.059\ m_0$. For the calculation of $\mathcal{A}$ we can assume $\omega \simeq \mathcal{E}_g/\hbar = 2\pi c/\lambda_0 = 2.16 \times 10^{15}$ rad/s, where $\lambda_0 = 870$ nm, so that

$$\mathcal{A} = 1.12 \times 10^4\ \text{cm}^{-1}\text{eV}^{-1/2}$$

Figure 5.1 shows the calculated absorption coefficient as a function of energy: when the photon energy exceeds the energy gap of just 50 meV, the absorption coefficient is already of the order of 2500 cm^{-1}. ■

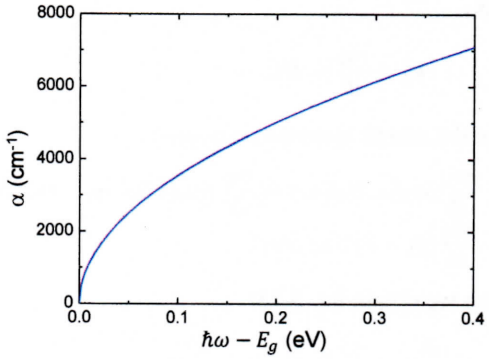

Figure 5.1: Plot of the absorption coefficient as a function of the difference between the photon energy, $\hbar\omega$, and the energy gap, $\mathcal{E}_g$, for an intrinsic GaAs bulk semiconductor calculated by using Eq. 5.38.

5.3.1 Direct calculation of absorption coefficient

It is now instructive to derive the expression of the absorption coefficient by using a different approach. Since the net number of absorption transitions is given by Eq. 5.25, the absorbed electromagnetic power is given by:

$$dP_{abs} = W_{12}\hbar\omega_0 dN = W_{12}\hbar\omega_0\rho_j(\omega_0)[f_v(\hbar\omega_0) - f_c(\hbar\omega_0)]d\omega_0 \tag{5.52}$$

where W_{12} is the transition rate corresponding to the transition with frequency ω_0. In other words, dP_{abs} is the electromagnetic power absorbed per unit volume when the transition frequency is in the range between ω_0 and $\omega_0 + d\omega_0$. Since the absorption coefficient is given by the ratio between the absorbed power per unit volume and the input intensity $I = \epsilon_0 ncE_0^2/2$, we have:

$$d\alpha = \frac{2dP_{abs}}{\epsilon_0 ncE_0^2} = \frac{2\hbar\omega_0}{\epsilon_0 ncE_0^2}W_{12}\rho_j(\omega_0)[f_v(\mathcal{E}_1) - f_c(\mathcal{E}_2)]d\omega_0 \tag{5.53}$$

where the transition rate W_{12} is given by

$$W_{12} = \frac{2\pi}{\hbar^2}|H'_{21}|^2 g(\omega - \omega_0)$$

The absorption coefficient can be calculated from Eq. 5.52 by integrating over all possible transition frequencies ω_0:

$$\alpha(\omega) = \int_{\mathcal{E}_g/\hbar}^{\infty} \frac{\omega_0\pi\mu_{21}^2}{nc\epsilon_0\hbar}\rho_j(\omega_0)g(\omega - \omega_0)[f_v(\hbar\omega_0) - f_c(\hbar\omega_0)]d\omega_0 \tag{5.54}$$

where we have used Eq. 5.44. If $g(\omega - \omega_0)$ is replaced by the delta-function $\delta(\omega - \omega_0)$ we find the same result obtained by using the approach based on the optical susceptibility (see Eq. 5.30).

5.4 Gain coefficient and Bernard-Duraffourg condition

The gain coefficient is given by $g(\omega) = -\alpha(\omega)$, therefore:

$$g(\omega) = \alpha_0(\omega)[f_c(\hbar\omega) - f_v(\hbar\omega)] \tag{5.55}$$

Therefore, the condition for optical gain is given by:

$$f_c(\hbar\omega) - f_v(\hbar\omega) > 0 \tag{5.56}$$

that is:

$$1 + \exp\left(\frac{\mathcal{E}_2 - \mathcal{E}_{Fc}}{k_B T}\right) < 1 + \exp\left(\frac{\mathcal{E}_1 - \mathcal{E}_{Fv}}{k_B T}\right), \tag{5.57}$$

which implies:

$$\mathcal{E}_2 - \mathcal{E}_1 < \mathcal{E}_{Fc} - \mathcal{E}_{Fv} \tag{5.58}$$

which represents the *Bernard-Duraffourg condition*: only those photons with energy $\hbar\omega = \mathcal{E}_2 - \mathcal{E}_1$ smaller than the energy separation between the two quasi-Fermi levels present a positive gain and are amplified. Since $\hbar\omega = \mathcal{E}_2 - \mathcal{E}_1 > \mathcal{E}_g$, in order to have net gain we must have:

$$\mathcal{E}_g < \hbar\omega < \mathcal{E}_{Fc} - \mathcal{E}_{Fv} \tag{5.59}$$

The limiting case, $\mathcal{E}_{Fc} - \mathcal{E}_{Fv} = \mathcal{E}_g$, defines the *transparency condition*. In this case $g = 0$ at $\omega = \mathcal{E}_g/\hbar$. To reach the transparency condition electrons must be injected into the conduction band: the corresponding density is called transparency density, N_{tr}. In the case of GaAs $N_{tr} \simeq 10^{18}$ cm^{-3} . When the density of the electrons injected into the conduction band exceeds the transparency density, $\mathcal{E}_{Fc} - \mathcal{E}_{Fv} > \mathcal{E}_g$ and net gain is present in the photon energy range between $\mathcal{E}_g$ and $\mathcal{E}_{Fc} - \mathcal{E}_{Fv}$. Figure 5.2 shows the gain coefficient $g(\omega)$ in the case of GaAs, for different values of the injected electron density. As shown in Fig. 5.2, even at high injection levels, the gain

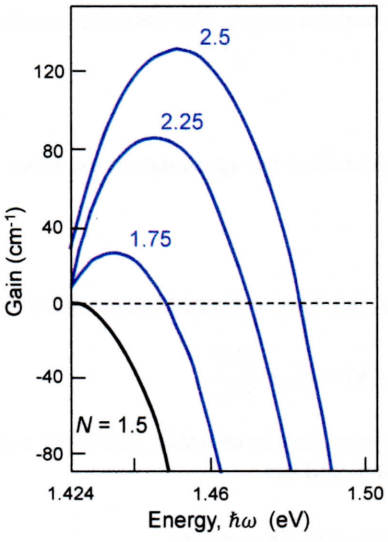

Figure 5.2: Gain coefficient as a function of photon energy for different values of the injected carrier density, N, (in units of 10^{18} cm^{-3}) (adapted from A. Yariv, *Quantum Electronics*, (Wiley, New York, 1989)).

bandwidth is only a small fraction of the energy gap. The peak of the gain curve

increases upon increasing the density of the injected electrons. For typical gain values ($20 \leq g \leq 80$ cm^{-1}) the gain peak increases almost linearly with N:

$$g = \sigma(N - N_{tr}) \tag{5.60}$$

where σ is called *differential gain coefficient* ($\sigma \simeq 1.5 \times 10^{-16}$ cm^2 in the case of GaAs).

5.5 Spontaneous emission

In this section we will consider the process of spontaneous emission, without entering into too many details. The semiclassical approach adopted so far, in which the radiation is treated classically (i.e., by using the Maxwell's equations), while matter is quantized and therefore treated using quantum mechanics, cannot be used to analyse the process of spontaneous emission, which has to be treated by using quantum electrodynamics, in which both matter and radiation are quantized. A very powerful approach, which does not require field quantization, is based on Einstein thermodynamic treatment. By using the Einstein's coefficient A_{cv}, the spontaneous emission rate, $r_{sp}(\mathcal{E})$, for a given transition from the level $\mathcal{E}_2$ in the conduction band to the level $\mathcal{E}_1$ in the valence band, with the same wavevector $\mathbf{k}$, is independent of the photon density in the semiconductor and can be written as follows:

$$r_{sp}(\mathcal{E}) = A_{cv} f_c(\mathcal{E}_2)[1 - f_v(\mathcal{E}_1)] \tag{5.61}$$

where A_{cv} is the spontaneous transition rate in the volume V. $r_{sp}(\mathcal{E})$ is proportional to the probability that the state with energy $\mathcal{E}_2$ in the conduction band is filled by an electron and to the probability that the corresponding state $\mathcal{E}_1$ (i.e., the state with the same $\mathbf{k}$) in the valence band is empty. The Einstein's coefficient A_{cv} can be written as:

$$A_{cv} = 1/\tau_r, \tag{5.62}$$

where τ_r is the radiative lifetime (or spontaneous emission lifetime) given by:

$$\tau_r = \frac{\pi \epsilon_0 \hbar c^3}{e^2 x_{vc}^2 n \omega_{vc}^3} \tag{5.63}$$

where $\hbar\omega_{vc} = \mathcal{E}_c(\mathbf{k}) - \mathcal{E}_v(\mathbf{k})$. Due to energy conservation we must have:

$$\hbar\omega_{vc} = \mathcal{E}_c(\mathbf{k}) - \mathcal{E}_v(\mathbf{k}) = \mathcal{E}_g + \frac{\hbar^2 k^2}{2m_r} \tag{5.64}$$

If we now consider the joint density of states, $\rho_j(\mathcal{E})$, the total spontaneous emission rate $R_{sp}(\hbar\omega)$ can be calculated as:

$$R_{sp}(\hbar\omega) = \int r_{sp}(\mathcal{E}) \rho_j(\mathcal{E}) g(\mathcal{E} - \hbar\omega) d\mathcal{E} \tag{5.65}$$

$r_{sp}(\mathcal{E})\rho_j(\mathcal{E})$ is slowly varying compared to the lineshape function, therefore in Eq. 5.65 we can replace $g(\mathcal{E} - \hbar\omega)$ with $\delta(\mathcal{E} - \hbar\omega)$, so that Eq. 5.65 simplifies to

$$R_{sp}(\hbar\omega) = r_{sp}(\hbar\omega)\rho_j(\hbar\omega). \tag{5.66}$$

From Eqs. 5.66, 5.61 and 5.63 we can write:

$$R_{sp}(\hbar\omega) = \frac{1}{\tau_r}\rho_j(\hbar\omega) f_c(\hbar\omega)[1 - f_v(\hbar\omega)]. \tag{5.67}$$

We note that at thermal equilibrium with background thermal radiation, the ratio between the spontaneous emission rate, $R_{sp}(\hbar\omega)$, and the absorption $\alpha(\hbar\omega)$:

$$\alpha(\hbar\omega) = \frac{\omega\pi e^2 x_{cv}^2}{cn\epsilon_0}\rho_j(\hbar\omega)[f_v(\hbar\omega) - f_c(\hbar\omega)], \tag{5.68}$$

is given by the following relation, called *van Roosbroeck-Shockley equation*

$$\frac{R_{sp}(\hbar\omega)}{\alpha(\hbar\omega)} = \frac{(\hbar\omega)^2 n^2}{\pi^2 c^2 \hbar^3} \frac{1}{\exp(\hbar\omega/k_B T) - 1}. \tag{5.69}$$

The "non-equilibrium" van Roosbroeck-Shockley equation is given by:

$$\frac{R_{sp}(\hbar\omega)}{\alpha(\hbar\omega)} = \frac{(\hbar\omega)^2 n^2}{\pi^2 c^2 \hbar^3} \frac{f_c(\hbar\omega)[1 - f_v(\hbar\omega)]}{f_v(\hbar\omega) - f_c(\hbar\omega)}, \tag{5.70}$$

which reduces to Eq. 5.69 at thermal equilibrium (when $\mathcal{E}_{Fc} = \mathcal{E}_{Fv} = \mathcal{E}_F$).

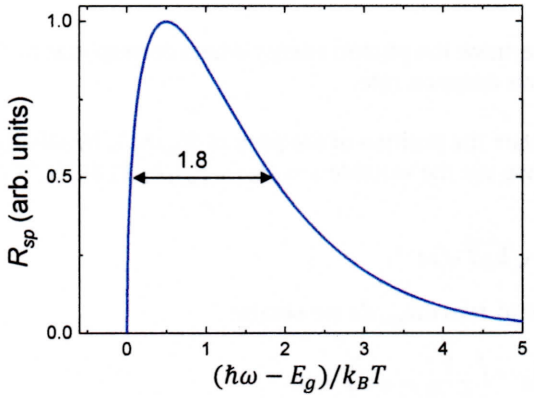

Figure 5.3: Spectral distribution of the spontaneous emission rate as a function of $(\hbar\omega - \mathcal{E}_g)/k_B T$, calculated by using Eq. 5.77.

If the Fermi-Dirac distribution functions can be approximated by the Boltzmann distribution function (i.e., if the quasi-Fermi levels are far from the bandedges), we can write:

$$f_c(\mathcal{E}_2) \approx \exp\left(-\frac{\mathcal{E}_2 - \mathcal{E}_{Fc}}{k_B T}\right), \tag{5.71}$$

$$1 - f_v(\mathcal{E}_1) \approx \exp\left(-\frac{\mathcal{E}_{Fv} - \mathcal{E}_1}{k_B T}\right), \tag{5.72}$$

so that:

$$f_c(\hbar\omega)[1 - f_v(\hbar\omega)] \approx \exp\left(-\frac{\hbar\omega}{k_B T} + \frac{\Delta\mathcal{E}_F}{k_B T}\right), \tag{5.73}$$

where $\Delta\mathcal{E}_F = \mathcal{E}_{Fc} - \mathcal{E}_{Fv}$. With these approximations $R_{sp}(\hbar\omega)$ can be written as follows:

$$R_{sp}(\hbar\omega) = \frac{1}{\tau_r}\rho_j(\hbar\omega)\exp\left(-\frac{\hbar\omega}{k_BT}\right)\exp\left(\frac{\Delta\mathcal{E}_F}{k_BT}\right) \tag{5.74}$$

where $\rho_j(\hbar\omega) = \rho_j(\mathcal{E})$ is given by Eq. 5.20 so that:

$$R_{sp}(\hbar\omega) = \frac{1}{\tau_r}\frac{1}{2\pi^2}\left(\frac{2m_r}{\hbar^2}\right)^{3/2}(\hbar\omega - \mathcal{E}_g)^{1/2}\exp\left(-\frac{\hbar\omega - \mathcal{E}_g}{k_BT}\right)\exp\left(\frac{\Delta\mathcal{E}_F - \mathcal{E}_g}{k_BT}\right). \tag{5.75}$$

If we now introduce the parameter $\Re_{sp}$ defined as

$$\Re_{sp} = \frac{1}{2\pi^2\tau_r}\left(\frac{2m_r}{\hbar^2}\right)^{3/2}\exp\left(\frac{\Delta\mathcal{E}_F - \mathcal{E}_g}{k_BT}\right) \tag{5.76}$$

Eq. 5.75 can be re-written as follows:

$$R_{sp}(\hbar\omega) = \Re_{sp}\,(\hbar\omega - \mathcal{E}_g)^{1/2}\exp\left(-\frac{\hbar\omega - \mathcal{E}_g}{k_BT}\right). \tag{5.77}$$

Figure 5.3 shows the spontaneous emission rate $R_{sp}(\hbar\omega)$ as a function of $(\hbar\omega - \mathcal{E}_g)/k_BT$.

Exercise 5.3 Determine the photon energy which corresponds to the peak value of the spontaneous emission rate.

We have to calculate the position of the peak of $R_{sp}(\hbar\omega)$, by solving the equation $dR_{sp}/d\omega = 0$. If we use the variable $z = (\hbar\omega - \mathcal{E}_g)/k_BT$, Eq. 5.77 can be written as

$$R_{sp} = \Re_{sp}\sqrt{k_BT}\sqrt{z}e^{-z}. \tag{5.78}$$

If we now equate to zero dR_{sp}/dz we obtain:

$$\Re_{sp}\sqrt{k_BT}e^{-z}\left(\frac{1}{2\sqrt{z}} - \sqrt{z}\right) = 0. \tag{5.79}$$

which gives $z = 1/2$. Therefore the photon energy corresponding to the spontaneous emission peak is given by:

$$\hbar\omega = \mathcal{E}_g + \frac{k_BT}{2}. \tag{5.80}$$

■

Exercise 5.4 Determine the full-width at half-maximum (FWHM) of the spontaneous emission spectrum. Assuming an emission wavelength $\lambda = 580$ nm, calculate the FWHM of the spontaneous emission spectrum (write the result in nanometers.)

The peak of the function $R_{sp}(\hbar\omega)$ can be calculated from Eq. 5.78 by using

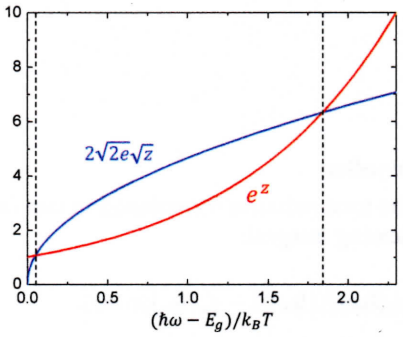

Figure 5.4: Graphical solution of Eq. 5.82.

$z = 1/2$:

$$R_{sp,max} = \Re_{sp} \sqrt{\frac{k_B T}{2e}} \tag{5.81}$$

where e is the Neper number. The FWHM can be calculated by solving the following equation:

$$2\sqrt{2e}\sqrt{z} = e^z \tag{5.82}$$

By using a simple graphical method (see Fig. 5.4) we obtain the FWHM:

$$\Delta\mathcal{E} \approx 1.8 k_B T. \tag{5.83}$$

Since $\Delta\mathcal{E} = h\Delta\nu$ and $\Delta\lambda/\lambda = \Delta\nu/\nu$, the FWHM of the spontaneous emission spectrum can be written in terms of the wavelength as follows:

$$\Delta\lambda \approx \frac{1.8\lambda^2 k_B T}{hc}. \tag{5.84}$$

For example, assuming that the wavelength corresponding to the maximum emission is $\lambda = 580$ nm at $T = 300$ K, the FWHM of the spectrum is given by $\Delta\lambda \simeq 12.6$ nm. ■

Exercise 5.5 Calculate the spontaneous lifetime of GaAs by using the numerical data reported in Ex. 5.2.

From Eq. 5.63 we have:

$$\tau_r = \frac{\pi\epsilon_0\hbar c^3}{e^2 x_{vc}^2 n\omega_{vc}^3} = \frac{\pi\epsilon_0\hbar^4 c^3}{e^5 x_{vc}^2 n(\hbar\omega_{vc}/e)^3},$$

where $\hbar\omega_{vc} \simeq \mathcal{E}_g$ so that $\hbar\omega_{vc}/e = 1.424$ eV. By substituting the numerical values

we obtain:

$$\tau_r \simeq 0.8 \text{ ns}.$$

■

5.5.1 Bimolecular recombination

We can now calculate the total radiative recombination rate for photons of all energies by considering the following integral:

$$R_{sp,T} = \frac{1}{\tau_r} \int_{\mathcal{E}_g}^{\infty} \rho_j(\hbar\omega) f_c(\hbar\omega)[1 - f_v(\hbar\omega)] d(\hbar\omega). \tag{5.85}$$

By using Eq. 5.74 we can write:

$$R_{sp,T} = \frac{e^{\Delta\mathcal{E}_F/k_BT}}{\tau_r} \int_{\mathcal{E}_g}^{\infty} \rho_j(\hbar\omega) e^{-\hbar\omega/k_BT} d(\hbar\omega). \tag{5.86}$$

If we now use the expression of $\rho_j(\hbar\omega) = \rho_j(\mathcal{E})$ given by Eq. 5.20 we obtain:

$$\begin{aligned} R_{sp,T} &= \frac{e^{\Delta\mathcal{E}_F/k_BT}}{\tau_r} \frac{1}{2\pi^2} \left(\frac{2m_r}{\hbar^2}\right)^{3/2} \int_{\mathcal{E}_g}^{\infty} (\hbar\omega - \mathcal{E}_g)^{1/2} e^{-\hbar\omega/k_BT} d(\hbar\omega) = \\ &= \frac{e^{(\Delta\mathcal{E}_F - \mathcal{E}_g)/k_BT}}{\tau_r} \frac{1}{2\pi^2} \left(\frac{2m_r k_BT}{\hbar^2}\right)^{3/2} \int_0^{\infty} \frac{\sqrt{x}}{e^x} dx, \end{aligned} \tag{5.87}$$

where $x = (\hbar\omega - \mathcal{E}_g)/k_BT$. Since:

$$\int_0^{\infty} \frac{\sqrt{x}}{e^x} dx = \frac{\sqrt{\pi}}{2}, \tag{5.88}$$

from Eq. 5.87 one obtains:

$$R_{sp,T} = \frac{1}{\tau_r} N_r \exp\left(\frac{\Delta\mathcal{E}_F - \mathcal{E}_g}{k_BT}\right), \tag{5.89}$$

where, in analogy with the effective density of states N_c and N_v introduced in Ch. 2, we have used the effective density of states N_r defined as:

$$N_r = \frac{1}{4} \left(\frac{2m_r k_BT}{\pi\hbar^2}\right)^{3/2}. \tag{5.90}$$

Since $n = N_c e^{-\frac{\mathcal{E}_c - \mathcal{E}_{Fc}}{k_BT}}$ and $p = N_v e^{-\frac{\mathcal{E}_{Fv} - \mathcal{E}_v}{k_BT}}$, Eq. 5.89 can be re-written as follows:

$$R_{sp,T} = \frac{1}{\tau_r} \frac{N_r}{N_c N_v} np = Bnp, \tag{5.91}$$

where B is called *bimolecular recombination coefficient*

$$B = \frac{1}{\tau_r} \frac{N_r}{N_c N_v} = \frac{1}{\tau_r N_c} \left(\frac{m_r}{m_v}\right)^{3/2}, \tag{5.92}$$

with typical values in the range $10^{-11} - 10^{-9}$ cm^3/s for III-V semiconductors.

	B (cm^3s^{-1})
GaAs	1.7×10^{-10}
AlAs	7.5×10^{-11}
InP	1.2×10^{-10}
AlGaN	2.0×10^{-10}

Table 5.2: Bimolecular recombination coefficients for various direct gap semiconductors at 300 K.

At thermal equilibrium the recombination rate is equal to the thermal generation rate, G_0, so that we can write:

$$R_0 = G_0 = Bn_0p_0 = Bn_i^2, \tag{5.93}$$

where n_0 and p_0 are the equilibrium carrier densities and we have used the mass-action law $n_0p_0 = n_i^2$. In the presence of excess electron-hole pairs $\Delta n = \Delta p$ (to maintain charge neutrality), the non-equilibrium carrier densities are given by:

$$n = n_0 + \Delta n, \qquad p = p_0 + \Delta p. \tag{5.94}$$

These excess carriers can be generated by absorption of light or by an injection current. In this situation the net radiative recombination rate is given by:

$$R_r = Bnp - R_0 = B(np - n_0p_0) \tag{5.95}$$

When carriers are injected into the semiconductor, the product np is larger than in the case of thermodynamic equilibrium, so that $np > n_0p_0$ and the recombination rate is positive.
The recombination dynamics can be described by the following equation:

$$\frac{dn}{dt} = \frac{dp}{dt} = R_0 - Bnp = -B(np - n_0p_0) = -B(np - n_i^2) \tag{5.96}$$

where $dn/dt = d\Delta n/dt$ and $dp/dt = d\Delta p/dt$ (moreover we assume $\Delta n = \Delta p$. Equation 5.96 can be re-written as:

$$\frac{d\Delta n}{dt} = -B(n_0\Delta p + p_0\Delta n + \Delta n\Delta p). \tag{5.97}$$

The general solution of Eq. 5.97 is given by:

$$\Delta n = \frac{(n_0 + p_0)\Delta n(0)}{[n_0 + p_0 + \Delta n(0)]\exp[B(n_0 + p_0)t] - \Delta n(0)}, \tag{5.98}$$

where $\Delta n(0)$ is the excess electron density at $t = 0$. In the case of low injection $\Delta n = \Delta p \ll n_0, p_0$, so that Eq. 5.97 can be written as:

$$\frac{d\Delta n}{dt} = -B(n_0 + p_0)\Delta n, \tag{5.99}$$

whose solution is:

$$\Delta n = \Delta n(0)\exp[-B(n_0 + p_0)t] = \Delta n(0)\exp\left(-\frac{t}{\tau_{rad}}\right), \tag{5.100}$$

where

$$\tau_{rad} = \frac{1}{B(n_0 + p_0)} \tag{5.101}$$

is the radiative recombination time. In the case of low injection the net radiative recombination rate of Eq. 5.95 can be written as:

$$R_r = B(np - n_0 p_0) \simeq B(n_0 + p_0)\Delta n, \tag{5.102}$$

which can be rewritten by using the definition of τ_{rad}:

$$R_r = \frac{\Delta n}{\tau_{rad}}. \tag{5.103}$$

In a $n-$type semiconductor $n_0 \gg p_0$ so that the radiative recombination time can be written as:

$$\tau_{rad} = \frac{1}{Bn_0} = \frac{1}{BN_D}, \tag{5.104}$$

where N_D is the donor density. In the same way, in a $p-$type semiconductor $p_0 \gg n_0$ and the radiative recombination time is given by:

$$\tau_{rad} = \frac{1}{Bp_0} = \frac{1}{BN_A}, \tag{5.105}$$

where N_A is the acceptor density.
In the case of strong injection, $n \approx p \gg n_0, p_0$ and Eq. 5.97 can be written as follows:

$$\frac{d\Delta n}{dt} = -B(\Delta n)^2, \tag{5.106}$$

and the solution is:

$$\Delta n(t) = \frac{\Delta n(0)}{1 + B\, t\, \Delta n(0)}. \tag{5.107}$$

Exercise 5.6 Consider a GaAs bulk semiconductor at $T = 300$ K $n-$doped with Silicon ($N_D = 10^{17}$ cm^{-3}); the spontaneous emission lifetime is $\tau_r = 0.8$ ns. Calculate the bimolecular recombination coefficient and the radiative recombination time, τ_{rad}, assuming low injection levels.

In the case of low-injection in an $n-$type semiconductor, the radiative lifetime is given by Eq. 5.104, where the bimolecular recombination coefficient is given by Eq. 5.92. We have first to calculate the effective density of states N_c:

$$N_c = \frac{1}{4}\left(\frac{2m_c k_B T}{\pi\hbar^2}\right)^{3/2} = 4.4 \times 10^{17}\ \text{cm}^{-3}.$$

Since the reduced mass is $m_r = 0.059\, m_0$, the bimolecular recombination coefficient is given by:

$$B = \frac{1}{\tau_r N_c}\left(\frac{m_r}{m_v}\right)^{3/2} = 1.3 \times 10^{-10}\ \text{cm}^3\text{s}^{-1}$$

and

$$\tau_{rad} = \frac{1}{BN_D} = 77 \text{ ns}.$$

5.6 Nonradiative recombination

There are processes which lead to electron-hole recombination without photon emission. In these cases the electron energy is converted to vibrational energy of the lattice, i.e., the recombination generates phonons instead of photons. Nonradiative recombination can be determined by the presence of defects in the semiconductor, which generally lead to the formation of electronic energy levels within the bandgap of the semiconductor. Other origins of nonradiative recombination are the Auger process and the recombination at surface.

5.6.1 Shockley-Read-Hall model

Energy levels within the gap of semiconductors generated by the presence of impurities are efficient recombination centers, which are particularly effective when they are close to the middle of the bandgap. In the framework of the *Shockley-Read-Hall (SRH) model*, four processes involving trap states are possible: i) an electron from the conduction band can be trapped in an energy trap; (ii) a filled trap can emit an electron in the conduction band; (iii) an empty trap can emit a hole in the valence band, or, in other words, an electron from the valence band can be trapped in an empty trap; (iv) a filled trap can emit an electron in the valence band. These processes are schematically shown in Fig. 5.5. In the following the essential ingredients

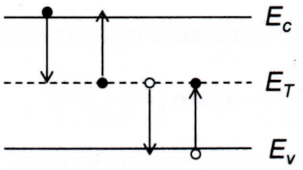

Figure 5.5: Nonradiative processes involving trap states.

of the SRH model will be outlined.

Let us assume that trap states are present in the gap at energy $\mathcal{E}_T$, with a density N_T (number of traps per unit volume). The density of filled traps is given by:

$$N_T f(\mathcal{E}_T) \tag{5.108}$$

where $f(\mathcal{E}_T)$ is the distribution function giving the probability of occupation of the trap state. The density of empty traps is

$$N_T[1 - f(\mathcal{E}_T)] \tag{5.109}$$

It is evident that the capture rate of electrons in the trap states is proportional to the product of the number of empty traps and of the electron density in the conduction band. It is generally assumed that the proportionality constant can be written as $C_n = v_{th}\sigma_n$, where v_{th} is the thermal velocity (see Ex. 5.1) $v_{th} = (3k_BT/m^*)^{1/2}$

($v_{th} \simeq 10^7$ cm/s at room temperature) and σ_n is called electron capture cross section, which turns out to be of the order of the square of the lattice constant and is typically $\sigma_n \simeq 10^{-15}$ cm^2. Therefore, the capture rate for electrons can be written as follows:

$$R_{cn} = C_n n N_T[1 - f(\mathcal{E}_T)] = v_{th}\sigma_n n N_T[1 - f(\mathcal{E}_T)]. \tag{5.110}$$

The emission rate from filled traps is proportional to the number of filled traps an can be written as

$$R_{en} = e_n N_T f(\mathcal{E}_T). \tag{5.111}$$

If the semiconductor is in thermodynamic equilibrium, the capture rate for electrons must equal the emission rate for electrons so that

$$R^0_{cn} = R^0_{en} \tag{5.112}$$

thus giving:

$$e_n = C_n n_0 \frac{1 - f_0(\mathcal{E}_T)}{f_0(\mathcal{E}_T)}, \tag{5.113}$$

where n_0 is the equilibrium electron density and $f_0(\mathcal{E}_T)$ is the equilibrium distribution function:

$$f_0(\mathcal{E}_T) = \frac{1}{1 + e^{\frac{\mathcal{E}_T - \mathcal{E}_F}{k_B T}}} \tag{5.114}$$

Therefore Eq. 5.113 can be rewritten as:

$$e_n = C_n n_0 \exp\left(\frac{\mathcal{E}_T - \mathcal{E}_F}{k_B T}\right) \tag{5.115}$$

From Eqs. 2.47 and 2.55 the equilibrium electron density n_0 can be written as:

$$n_0 = n_i \exp\left(\frac{\mathcal{E}_F - \mathcal{E}_i}{k_B T}\right) = N_c \exp\left(\frac{\mathcal{E}_F - \mathcal{E}_c}{k_B T}\right) \tag{5.116}$$

where $\mathcal{E}_i$ is the intrinsic Fermi level, equal to the Fermi level in the intrinsic semiconductor. By using Eq. 5.116, Eq. 5.115 can be written as:

$$e_n = C_n n_i \exp\left(\frac{\mathcal{E}_T - \mathcal{E}_i}{k_B T}\right) = C_n N_c \exp\left(\frac{\mathcal{E}_T - \mathcal{E}_c}{k_B T}\right) = C_n n_1 \tag{5.117}$$

where:

$$n_1 = n_i \exp\left(\frac{\mathcal{E}_T - \mathcal{E}_i}{k_B T}\right) = N_c \exp\left(\frac{\mathcal{E}_T - \mathcal{E}_c}{k_B T}\right) \tag{5.118}$$

If the semiconductor is not in thermodynamic equilibrium, capture rates and emission rates are not equal. In this case the net trap capture rate is given by the difference between the capture and emission rates:

$$R_n = R_{cn} - R_{en}. \tag{5.119}$$

By using Eqs. 5.110-5.111 and Eq. 5.117, R_n can be written as:

$$R_n = C_n N_T\{n[1 - f(\mathcal{E}_T)] - n_1 f(\mathcal{E}_T)\}, \tag{5.120}$$

where $f(\mathcal{E}_T)$ is a non-equilibrium distribution function, whose expression will be calculated below.

We can repeat the same discussion for the derivation of the net hole recombination rate R_p, given by the difference between the hole capture and emission rates. The hole capture rate must be proportional to the hole density and to the number of filled traps:

$$R_{cp} = C_p p N_T f(\mathcal{E}_T) = v_{th}\sigma_p p N_T f(\mathcal{E}_T), \tag{5.121}$$

where we have used the cross-section for hole capture, σ_p. The hole emission rate is related to the number of empty traps: an electron in the valence band is captured by an empty trap, thus leaving a hole in the valence band. Since there are plenty of electrons in the valence band, the hole emission is not limited by the electron concentration in the valence band. Therefore:

$$R_{ep} = e_p N_T[1 - f(\mathcal{E}_T)]. \tag{5.122}$$

By following the same procedure used in the case of electrons, the parameter e_p can be written as follows:

$$e_p = C_p p_1, \tag{5.123}$$

where p_1 is given by:

$$p_1 = n_i \exp\left(-\frac{\mathcal{E}_T - \mathcal{E}_i}{k_B T}\right) = N_v \exp\left(-\frac{\mathcal{E}_T - \mathcal{E}_v}{k_B T}\right). \tag{5.124}$$

Therefore, the net hole recombination rate R_p is:

$$R_p = R_{cp} - R_{ep} = C_p N_T\{p f(\mathcal{E}_T) - p_1[1 - f(\mathcal{E}_T)]\}. \tag{5.125}$$

In steady-state the net electron capture rate must be equal to the net hole recombination rate, $R_n = R_p$. By equating Eqs. 5.120 and 5.125 the non-equilibrium distribution function $f(\mathcal{E}_T)$ can be obtained:

$$f(\mathcal{E}_T) = \frac{C_n n + C_p p_1}{C_n(n + n_1) + C_p(p + p_1)}. \tag{5.126}$$

We note that from Eqs. 5.118 and 5.124 the product $n_1 p_1$ is simply given by:

$$n_1 p_1 = n_i^2 \tag{5.127}$$

therefore, the recombination rates $R_n = R_p$ can be written as follows:

$$R_n = R_p = \frac{C_n C_p N_T(np - n_i^2)}{C_n(n + n_1) + C_p(p + p_1)}. \tag{5.128}$$

From the previous equation it is evident that when the semiconductor is in thermodynamic equilibrium $np = n_i^2$ so that $R_n = R_p = 0$, as we have already pointed out. If there is an excess carrier density in the semiconductor, np turns out to be larger than n_i^2 and the system reacts trying to establish an equilibrium condition by exploiting recombination processes: indeed Eq. 5.128 gives a positive recombination rate. In the opposite case, when there is a net carrier depletion, $np < n_i^2$ so that $R_n = R_p < 0$: in this case we have a net generation rate.

It is common practice to introduce the electron and hole lifetimes, τ_n and τ_p, respectively, defined by the following expressions:

$$\tau_n = \frac{1}{C_n N_T}, \qquad \tau_p = \frac{1}{C_p N_T}. \tag{5.129}$$

In terms of τ_n and τ_p the recombination rates can be written as:

$$R_n = R_p = \frac{np - n_i^2}{\tau_p(n + n_1) + \tau_n(p + p_1)}. \tag{5.130}$$

Now, the non-equilibrium electron and hole concentrations are given by:

$$n = n_0 + \Delta n, \qquad p = p_0 + \Delta p \tag{5.131}$$

so that the recombination rate can be rewritten as:

$$R_n = R_p = \frac{(n_0 + \Delta n)(p_o + \Delta p) - n_i^2}{\tau_p(n_0 + \Delta n + n_1) + \tau_n(p_0 + \Delta p + p_1)}. \tag{5.132}$$

Let us now consider a $n-$type semiconductor, so that $n_0 \gg p_0$, moreover we will assume that the excess carrier concentration is much smaller than the equilibrium electron concentration $\Delta n \ll n_0$. Moreover, if we further assume that the trap levels are close to the intrinsic Fermi level, $\mathcal{E}_T \approx \mathcal{E}_i$, we have:

$$n_1 \simeq n_i, \qquad p_1 \simeq n_i. \tag{5.133}$$

With these assumptions Eq. 5.132 becomes:

$$R = \frac{n_0 \Delta p}{\tau_p(n_0 + \Delta n + n_i)}, \tag{5.134}$$

indeed, in the denominator of the ratio in Eq. 5.132 the second term is much smaller than the first one (all terms are much smaller than n_0) and it can be neglected. In Eq. 5.134 we can also neglect Δn and n_i with respect to n_0 so that:

$$R = \frac{\Delta p}{\tau_p}, \tag{5.135}$$

which gives the recombination rate in terms of the excess minority carrier concentration Δp and lifetime τ_p.

By using Eq. 5.128 it is simple to deduce the position of the trap levels within the bandgap which maximizes the recombination rate R. The energy $\mathcal{E}_T$ can be calculated by equating to zero the derivative of R with respect to $\mathcal{E}_T$:

$$\frac{\partial R}{\partial \mathcal{E}_T} = 0 \tag{5.136}$$

so that

$$C_n \frac{dn_1}{d\mathcal{E}_T} + C_p \frac{dp_1}{d\mathcal{E}_T} = 0. \tag{5.137}$$

From Eqs. 5.118 and 5.124 we obtain:

$$\mathcal{E}_T = \frac{\mathcal{E}_c + \mathcal{E}_v}{2} + \frac{k_B T}{2} \log \frac{C_p N_v}{C_n N_c} \tag{5.138}$$

so that we can conclude that a recombination center is most effective when it is located close to the middle of the bandgap.

5.6.2 Auger recombination

Another important nonradiative recombination mechanism is related to the Auger process, in which the energy released by an electron-hole recombination event is not released in the form of photons but it is transferred to another electron or hole. The energy is eventually transferred in a nonradiative way from this third particle to the semiconductor lattice by phonon emission. In the direct band-to-band Auger process three particles are involved, thus it is a relevant recombination mechanism at high carrier concentration. This is generally the case of small band gap semiconductors or in systems which are far from thermodynamic equilibrium, for example in semiconductor lasers. Without entering into too many details, there are various possibilities, schematically shown in Fig. 5.6.

i) Electron capture. An electron from the conduction band recombines with a hole in

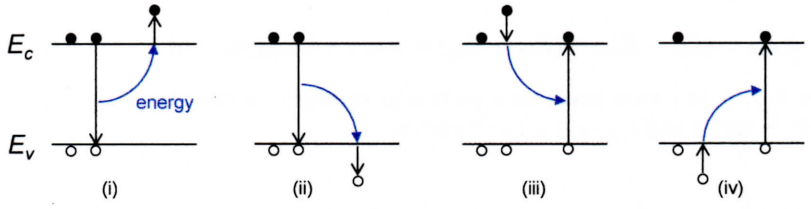

Figure 5.6: Various band-to-band Auger processes in a semiconductor (see text).

the valence band and the excess energy is transferred to an electron in the conduction band.

ii) Hole capture. Also in this second process an electron from the conduction band recombines with a hole in the valence band, but the excess energy is now transferred to an additional hole in the valence band.

iii) Electron emission. A high-energy electron in the conduction band transfers its energy to an electron in the valence band, which is thus excited in the conduction band leading to the generation of an electron-hole pair.

iv) Hole emission. A high-energy hole in the valence band transfers its energy to an electron in the valence band, which is excited into the conduction band thus generating an electron-hole pair.

A simple theory describing the Auger process can be introduced by using arguments similar to those introduced in the previous section in the case of the SRH model. We can first associate to the four processes described above a suitable rate. In particular, in the case of electron capture, two electrons in the conduction band and one hole in the valence band are required, so that an electron capture rate R_{ec} can be written as

$$R_{ec} = C_{ec} n^2 p, \tag{5.139}$$

where C_{ec} is a suitable proportionality constant. In a similar way, in the case of hole capture, two holes in the valence band and one electron in the conduction band are required, so that a hole capture rate R_{hc} can be written:

$$R_{hc} = C_{hc} n p^2. \tag{5.140}$$

In the case of electron emission only one electron in the conduction band is required so that:

$$R_{ee} = C_{ee} n \tag{5.141}$$

in the same way, for hole emission we can write:

$$R_{he} = C_{he}p. \tag{5.142}$$

In overall thermodynamic equilibrium the electron capture rate must be equal to the electron emission rate and the same for hole capture and emission rates, so that:

$$R_{ec}^0 = R_{ee}^0, \qquad R_{hc}^0 = R_{he}^0. \tag{5.143}$$

Therefore, in thermodynamic equilibrium we have:

$$C_{ec}n_i^2 = C_{ee}, \qquad C_{hc}n_i^2 = C_{he}. \tag{5.144}$$

If the semiconductor is not in thermodynamic equilibrium, the net Auger recombination rate is given by:

$$R_{Aug} = (R_{ec} - R_{ee}) + (R_{hc} - R_{he}) = (C_{ec}n + C_{hc}p)(np - n_i^2), \tag{5.145}$$

where Eqs. 5.144 were used. As a particular example, in the case of GaAs, $C_{ec} = 5 \times 10^{-30}$ cm^6/s and $C_{hc} = 3 \times 10^{-30}$ cm^6/s.

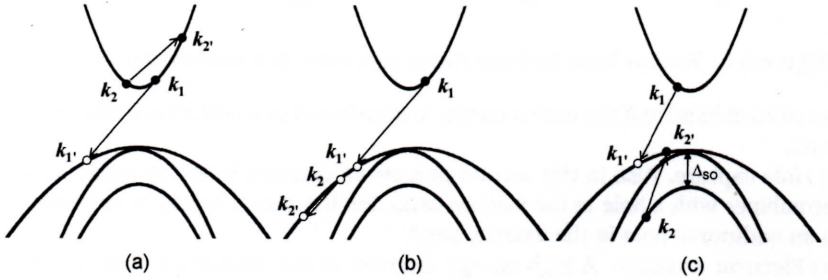

Figure 5.7: Three possible band-to-band direct Auger recombination processes: (a) CHCC, (b) CHHH and (c) CHSH.

Figure 5.7 schematically shows three possible Auger recombination processes and illustrates the energy and momentum conservation rules. In Fig. 5.7(a) an electron with wavevector $\mathbf{k}_1$ recombines with a hole with wavevector $\mathbf{k}_{1'}$ and the released energy is transferred to a second electron with initial wavevector $\mathbf{k}_2$, which is excited to a state with wavevector $\mathbf{k}_{2'}$. This process is called Conduction-Heavy hole-Conduction-Conduction (CHCC). Fig. 5.7(b) shows a process where an electron with wavevector $\mathbf{k}_1$ recombines with a hole with wavevector $\mathbf{k}_{1'}$ and the released energy is transferred to a second hole in the valence band with an initial wavevector $\mathbf{k}_2$ an final wavevector $\mathbf{k}_{2'}$. This process is called CHHH. Other Auger recombination processes are possible, involving also the light-hole valence band (CHLH processes) and the split-off band (CHSH processes). In Fig. 5.7(c) an electron with wavevector $\mathbf{k}_1$ recombines with a hole with wavevector $\mathbf{k}_{1'}$ and the released energy is transferred to an electron in the split-off band with wavevector $\mathbf{k}_2$, which is excited to a state in the heavy-hole valence band with wavevector $\mathbf{k}_{2'}$. In this case the energy gap of the semiconductor is larger than the split-off energy: $\mathcal{E}_g > \Delta_{SO}$.

Figure 5.7 refers to the simplest Auger process: the phononless band-to-band Auger process in a direct semiconductor. In this case the energy and momentum

conservation must be satisfied by the electrons and holes involved in the process. For this reason it is not possible to have a recombination between an electron at the bottom of the conduction band and a hole at the top of the valence band ($\Delta k = 0, \Delta\mathcal{E} = \mathcal{E}_g$) since the excited third particle cannot make a vertical transition. Therefore there must be an activation energy, which is related to the size of the energy gap. In the case of direct bandgap semiconductors with large gap this activation energy is large, thus leading to a decrease of the process probability. The activation energy is equal to the energy required to place the participating carriers in the appropriate threshold states for an Auger transition to occur.

The dependence of the band-to-band Auger process on the band structure, temperature and carrier density can be understood by using a relatively simple model. Let us assume that the dispersion relations $\mathcal{E}(\mathbf{k})$ can be well approximated by parabolic expressions both in the conduction and in the valence band and that the semiconductor is non-degenerate. The Fermi-Dirac distribution function can be approximated by the Boltzmann distribution function, so that the probability of occupation of a state in the conduction band can be written as:

$$f_c(\mathbf{k}) = \exp\left(-\frac{\mathcal{E}_c(\mathbf{k}) - \mathcal{E}_{Fc}}{k_B T}\right) \tag{5.146}$$

while the probability that a valence state is not occupied can be written as:

$$\bar{f}_v(\mathbf{k}) = \exp\left(-\frac{\mathcal{E}_{Fv} - \mathcal{E}_v(\mathbf{k})}{k_B T}\right) \tag{5.147}$$

where $\mathcal{E}_{Fc}$ and $\mathcal{E}_{Fv}$ are the quasi-Fermi energies in the conduction and valence band, respectively. $\mathcal{E}_c(\mathbf{k})$ and $\mathcal{E}_v(\mathbf{k})$ are given by:

$$\mathcal{E}_c(k) = \mathcal{E}_c + \frac{\hbar^2 k^2}{2m_c}, \qquad \mathcal{E}_v(k) = \mathcal{E}_v - \frac{\hbar^2 k^2}{2m_v} \tag{5.148}$$

where $\mathcal{E}_c$ corresponds to the bottom of the conduction band and $\mathcal{E}_v$ to the top of the valence band. It is evident that the probability of a CHCC process is related to the probability that the states in the conduction band corresponding to the wavevectors $\mathbf{k}_1$ and $\mathbf{k}_2$ are each occupied by an electron and to the probability that the state with wavevector $\mathbf{k}_{1'}$ in the valence band is empty; the state in the conduction band with wavevector $\mathbf{k}_{2'}$ can be assumed empty. Therefore we can write a probability P for the CHCC process as:

$$\begin{aligned} P(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_{1'}) &= f_c(\mathbf{k}_1) f_c(\mathbf{k}_2) \bar{f}_v(\mathbf{k}_{1'}) = \\ &= \exp\left[-\frac{\mathcal{E}_c(k_1) - \mathcal{E}_{Fc} + \mathcal{E}_c(k_2) - \mathcal{E}_{Fc} + \mathcal{E}_{Fv} - \mathcal{E}_v(k_{1'})}{k_B T}\right]. \end{aligned} \tag{5.149}$$

By using Eq. 5.148 we can write Eq. 5.149 as follows:

$$\begin{aligned} P(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_{1'}) &= \exp\left(-2\frac{\mathcal{E}_c - \mathcal{E}_{Fc}}{k_B T}\right) \exp\left(-\frac{\mathcal{E}_{Fv} - \mathcal{E}_v}{k_B T}\right) \\ &\cdot \exp\left[-\frac{\hbar^2[k_1^2 + k_2^2 + (m_c/m_v)k_{1'}^2]}{2m_c k_B T}\right]. \end{aligned} \tag{5.150}$$

In non-degenerate semiconductors the electrons and hole concentrations can be calculated as:

$$\begin{aligned} n &= N_c \exp\left(-\frac{\mathcal{E}_c - \mathcal{E}_{Fc}}{k_B T}\right) \\ p &= N_v \exp\left(-\frac{\mathcal{E}_{Fv} - \mathcal{E}_v}{k_B T}\right) \end{aligned} \tag{5.151}$$

so that the probability $P(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_{1'})$ can be written as:

$$P(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_{1'}) = \frac{n^2 p}{N_c^2 N_v} \exp\left[-\frac{\hbar^2[k_1^2 + k_2^2 + (m_c/m_v)k_{1'}^2]}{2m_c k_B T}\right]. \tag{5.152}$$

The momentum conservation requires that

$$\mathbf{k}_1 + \mathbf{k}_2 = \mathbf{k}_{1'} + \mathbf{k}_{2'} \tag{5.153}$$

and the energy conservation gives:

$$\mathcal{E}_c(k_1) + \mathcal{E}_c(k_2) = \mathcal{E}_v(k_{1'}) + \mathcal{E}_c(k_{2'}) \tag{5.154}$$

which can be also written as:

$$\frac{\hbar^2}{2m_c}\left[k_1^2 + k_2^2 + \left(\frac{m_c}{m_v}\right)k_{1'}^2 - k_{2'}^2\right] + \mathcal{E}_g = 0 \tag{5.155}$$

where $\mathcal{E}_g = \mathcal{E}_c - \mathcal{E}_v$ is the energy gap.

The most probable process is therefore obtained by maximizing $P(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_{1'})$ with the requirement that momentum and energy are conserved. This can be obtained by using the method of Lagrange multipliers (see the following gray box). The values of the wavevectors $\mathbf{k}_1$,$\mathbf{k}_2$ and $\mathbf{k}_{1'}$ which correspond to the most probable process are given by:

$$k_2 = k_1, \qquad k_{1'} = -\frac{m_v}{m_c} k_1 \tag{5.156}$$

so that $k_{2'}$ is given by:

$$k_{2'} = k_1 + k_2 - k_{1'} = \left(2 + \frac{m_v}{m_c}\right) k_1. \tag{5.157}$$

On the other hand, the value k_1 can be obtained from the energy conservation (see Eq. 5.155):

$$\frac{\hbar^2 k_1^2}{2m_c} = \frac{m_c^2}{(m_c + m_v)(2m_c + m_v)} \mathcal{E}_g \tag{5.158}$$

so that the maximum probability is given by the following formula:

$$P_{max} = \frac{n^2 p}{N_c^2 N_v} \exp\left(-\frac{m_c}{m_c + m_v} \frac{\mathcal{E}_g}{k_B T}\right). \tag{5.159}$$

It is instructive to rewrite Eq. 5.159 by using Eqs. 5.151:

$$\begin{aligned} P_{max} &= \frac{n}{N_c} \frac{np}{N_c N_v} \exp\left(-\frac{m_c}{m_c + m_v} \frac{\mathcal{E}_g}{k_B T}\right) = \\ &= \frac{n}{N_c} \exp\left(\frac{\mathcal{E}_{Fc} - \mathcal{E}_{Fv}}{k_B T}\right) \exp\left(-\frac{\mathcal{E}_g}{k_B T}\right) \exp\left(-\frac{m_c}{m_c + m_v} \frac{\mathcal{E}_g}{k_B T}\right) = \\ &= \frac{n}{N_c} \exp\left(\frac{\mathcal{E}_{Fc} - \mathcal{E}_{Fv}}{k_B T}\right) \exp\left(-\frac{2m_c + m_v}{m_c + m_v} \frac{\mathcal{E}_g}{k_B T}\right). \end{aligned} \tag{5.160}$$

If the semiconductor is in thermodynamic equilibrium ($\mathcal{E}_{Fc} - \mathcal{E}_{Fv} = 0$) and $n = n_0$, we have:

$$P_{max} = \frac{n_0}{N_c} \exp\left(-\frac{2m_c + m_v}{m_c + m_v}\frac{\mathcal{E}_g}{k_B T}\right) \tag{5.161}$$

It is evident that a CHHH process can be analyzed using the same procedure. In this case the maximum probability can be written by using Eq. 5.161 replacing n_0 with p_0, m_c with m_v and N_c with N_v, so that:

$$P_{max} = \frac{p_0}{N_v} \exp\left(-\frac{m_c + 2m_v}{m_c + m_v}\frac{\mathcal{E}_g}{k_B T}\right) \tag{5.162}$$

As an additional example we can consider the CHSH process, which leads to the generation of a hot hole in the split-off band, schematically shown in Fig. 5.7(c). In this case the exponential factor in the expression of the process probability is given by:

$$P_{CHSH} \propto \exp\left(-\frac{m_{SO}}{m_c + 2m_v - m_{SO}}\frac{\mathcal{E}_g - \Delta_{SO}}{k_B T}\right) \tag{5.163}$$

where m_{SO} is the split-off effective mass and Δ_{SO} is the split-off energy.
It is therefore evident that the probability of Auger processes increases upon decreasing the energy gap. Moreover, due to the presence of the exponential factor in the Auger probability, the Auger process turns out to be very sensitive to the band structure (which directly determine the values of the effective masses) and to the temperature.

5.6.1 Calculation of the values of k_1, k_2 and $k_{1'}$ which maximize the probability of the CHCC Auger process.
We have to find the maximum value of the function $P(k_1, k_2, k_{1'})$, given by Eq. 5.152, subject to the constraints imposed by the momentum and energy conservation expressed by Eqs. 5.153 and 5.155. We can use the method of Lagrange undetermined multipliers. To maximize $P(k_1, k_2, k_{1'})$ we require:

$$dP = \frac{\partial P}{\partial k_1}dk_1 + \frac{\partial P}{\partial k_2}dk_2 + \frac{\partial P}{\partial k_{1'}}dk_{1'} = 0 \tag{5.164}$$

If dk_1, dk_2 and $dk_{1'}$ were independent, we could conclude that:

$$\frac{\partial P}{\partial k_1} = \frac{\partial P}{\partial k_2} = \frac{\partial P}{\partial k_{1'}} = 0$$

however, they are not independent but constrained because of the momentum and energy conservation, so that by differentiating Eqs. 5.153 and 5.155 we obtain:

$$dk_1 + dk_2 - dk_{1'} = 0 \tag{5.165}$$

$$\frac{\hbar^2}{m_c}k_1 dk_1 + \frac{\hbar^2}{m_c}k_2 dk_2 + \frac{\hbar^2}{m_c}\frac{m_c}{m_v}k_{1'}dk_{1'} = 0 \tag{5.166}$$

On the other hand, equating dP to zero one obtains:

$$\frac{\hbar^2 k_1}{m_c k_B T} P(k_1, k_2, k_{1'}) dk_1 + \frac{\hbar^2 k_2}{m_c k_B T} P(k_1, k_2, k_{1'}) dk_2 + \\ + \frac{m_c}{m_v} \frac{k_{1'}}{m_c k_B T} P(k_1, k_2, k_{1'}) dk_{1'} = 0 \tag{5.167}$$

Multiplying Eq. 5.165 by an as yet unknown number λ and Eq. 5.166 by μ (λ and μ are the so called Lagrange undetermined multipliers) and adding them to Eq. 5.167 we obtain an equation where dk_1, dk_2 and $dk_{1'}$ are independent and arbitrary, so that we must choose λ and μ such that:

$$-\frac{\hbar^2 k_1}{m_c k_B T} P + \lambda + \mu \frac{\hbar^2}{m_c} k_1 = 0 \tag{5.168}$$

$$-\frac{\hbar^2 k_2}{m_c k_B T} P + \lambda + \mu \frac{\hbar^2}{m_c} k_2 = 0 \tag{5.169}$$

$$-\frac{m_c}{m_v} \frac{\hbar^2 k_{1'}}{m_c k_B T} P - \lambda + \mu \frac{\hbar^2}{m_c} \frac{m_c}{m_v} k_{1'} = 0 \tag{5.170}$$

which can be rewritten as follows:

$$-\frac{\hbar^2 k_1}{m_c k_B T} (P - \mu k_B T) + \lambda = 0 \tag{5.171}$$

$$-\frac{\hbar^2 k_2}{m_c k_B T} (P - \mu k_B T) + \lambda = 0 \tag{5.172}$$

$$-\frac{m_c}{m_v} \frac{\hbar^2 k_{1'}}{m_c k_B T} (P - \mu k_B T) - \lambda = 0 \tag{5.173}$$

These equations, together with the constraints given by momentum and energy conservation, are sufficient to determine the five unknown, λ, μ and the values of k_1, k_2 and $k_{1'}$ at the stationary point. Indeed, by subtracting Eq. 5.172 from Eq. 5.171 we have:

$$-\frac{\hbar^2}{m_c k_B T} (P - \mu k_B T)(k_1 - k_2) = 0 \tag{5.174}$$

from which we obtain:

$$k_1 = k_2 \tag{5.175}$$

If we now sum Eq. 5.171 and Eq. 5.173 we obtain:

$$-\frac{\hbar^2}{m_c k_B T} (P - \mu k_B T) \left(k_1 + \frac{m_c}{m_v} k_{1'} \right) = 0 \tag{5.176}$$

which gives

$$k_{1'} = -\frac{m_v}{m_c} k_1 \tag{5.177}$$

5.6.3 Surface recombination

Recombination at surfaces and interfaces in semiconductor structures can be quite important and can limit the performances of the semiconductor devices. As a consequence of the sudden termination of the semiconductor lattice structure at the surface, dangling bonds are generated at the surface and defects are very likely to form at surfaces. For example, oxygen is always found on a semiconductor surface and can produce surface states. The surface states which are neutral when occupied by electrons and positively charged when they are not occupied are called donor states, while states which are negatively charged when occupied and neutral when empty are called acceptor states. On the other hand, the interface between semiconductor and external environment cannot be abrupt on the atomic scale: it consists of a layer of intermediate materials and impurities with a thickness of several atomic layers. In this region there are surface states which are very close to the bulk semiconductor and which remain in thermal equilibrium with the bulk also when the potential is changed. Such states are called *fast surface states* since the electrons which occupy these states reach the equilibrium configuration rather quickly (on the order of microseconds or less). In addition, there are surface states which are in the intermediate layer, relatively far from the bulk semiconductor, so that they can reach thermal equilibrium with bulk states in relatively longer times (which can be of the order of seconds or more). For this reason they are called *slow surface states*.

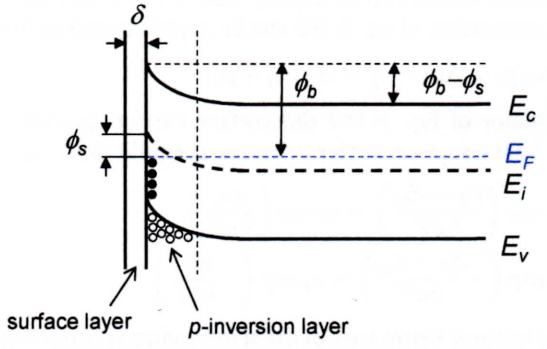

Figure 5.8: Schematic band diagram for a semiconductor showing a thin surface layer containing acceptor-type surface states.

The band structure of the semiconductor surface is strongly modified by the presence of surface states, as schematically shown in Fig. 5.8, where, to account for surface effects a surface layer with a thickness δ has been considered between the bulk semiconductor and the external environment (for example, vacuum). The figure refers to the case of an $n-$type semiconductor with a thin surface layer containing acceptor-type surface states: the semiconductor is depleted of electrons near the surface. It is also possible that the hole density at the surface, p_s, may become much larger than the electron density at the surface, n_s ($p_s \gg n_s$): in this case the surface becomes of $p-$type (this situation is shown in Fig. 5.8).

Surface states may become effective recombination centers. Surface recombination can be analyzed by using the same procedure employed in the case of bulk nonradiative recombination. In particular, within the SRH model the surface recom-

bination rate can be written by using Eq. 5.128, where n is repaced by the surface electron density n_s, p by the surface hole density p_s and N_T by the surface state density per unit area, N_{Ts}:

$$R_s = \frac{C_n C_p N_{Ts}(n_s p_s - n_i^2)}{C_n(n_s + n_1) + C_p(p_s + p_1)} \tag{5.178}$$

where, n_1 and p_1 are given by Eqs. 5.118 and 5.124, respectively ($\mathcal{E}_T$ in this case is the energy level of the fast surface states) and the electron and hole density at surface are related to the electron and hole density by:

$$n_s = n \exp\left[-\frac{q(\phi_b - \phi_s)}{k_B T}\right] \tag{5.179}$$

$$p_s = p \exp\left[\frac{q(\phi_b - \phi_s)}{k_B T}\right] \tag{5.180}$$

where ϕ_b and ϕ_s are reported in Fig. 5.8. Moreover:

$$n_s p_s = np = (n_0 + \Delta n)(p_0 + \Delta p) \tag{5.181}$$

where Δn and Δp are the excess electron and hole densities, respectively, so that:

$$R_s = \frac{C_n C_p N_{Ts}[(n_0 + \Delta n)(p_0 + \Delta p) - n_i^2]}{C_n(n_s + n_1) + C_p(p_s + p_1)} \tag{5.182}$$

In the case of small excess carrier density ($\Delta n = \Delta p \ll n_0$) the term in the square brackets at the numerator of Eq. 5.182 can be approximated as follows ($n_0 p_0 = n_i^2$):

$$(n_0 + \Delta n)(p_0 + \Delta p) - n_i^2 = \Delta n(n_0 + p_0). \tag{5.183}$$

In the denominator of Eq. 5.182 the surface carrier densities n_s and p_s can be approximated by the surface carrier densities at equilibrium, n_{s0} and p_{s0}, given by:

$$n_{s0} = n_i \exp\left(\frac{\mathcal{E}_F - \mathcal{E}_{is}}{k_B T}\right) = n_i \exp\left(\frac{q\phi_s}{k_B T}\right) \tag{5.184}$$

$$p_{s0} = n_i \exp\left(-\frac{\mathcal{E}_F - \mathcal{E}_{is}}{k_B T}\right) = n_i \exp\left(-\frac{q\phi_s}{k_B T}\right) \tag{5.185}$$

where $\mathcal{E}_{is}$ is the intrinsic Fermi level at the semiconductor surface and ϕ_s is the surface potential (see Fig. 5.8). With these approximations, the surface recombination velocity, defined as:

$$v_s = \frac{R_s}{\Delta n} = \frac{R_s}{\Delta p} \tag{5.186}$$

can be written as follows:

$$v_s = \frac{N_{Ts}\sqrt{C_n C_p}(n_0 + p_0)/2n_i}{\cosh[(\mathcal{E}_T - \mathcal{E}_i - q\phi_0)/k_B T] + \cosh[q(\phi_s - \phi_0)/k_B T]} \tag{5.187}$$

where:

$$q\phi_0 = \frac{k_B T}{2} \log \frac{C_p}{C_n} \tag{5.188}$$

Therefore the surface recombination velocity depends on the surface potential ϕ_s, which is very sensitive to the environment conditions at the surface: this is why in semiconductor devices it is important to maintain stable surface conditions. Moreover v_s is directly related to the surface state density.

5.7 Competition between radiative and nonradiative recombination

In a real semiconductor it is impossible to avoid completely nonradiative bulk and surface recombination and Auger recombination, which therefore will compete with the radiative recombination processes. The total recombination probability is given by the sum of the radiative and nonradiative recombination probabilities:

$$\frac{1}{\tau} = \frac{1}{\tau_{rad}} + \frac{1}{\tau_{nr}}, \tag{5.189}$$

so that the relative probability of radiative recombination, which is also called *internal quantum efficiency*, is given by:

$$\eta_{int} = \frac{\tau}{\tau_{rad}} = \frac{\tau_{nr}}{\tau_{rad} + \tau_{nr}}, \tag{5.190}$$

the internal quantum efficiency is the ratio between the number of photons emitted inside a semiconductor device and the total number of charge quanta, which undergo recombination.

6. Quantum Wells

6.1 Introduction

So far we have always considered bulk semiconductors, i.e., semiconductors with spatial dimensions much larger than the de Broglie wavelength of the electron involved in the relevant physical processes. In a quantum well (QW) a very thin layer of a semiconductor, with energy gap $\mathcal{E}_{g1}$, is grown between two semiconductors with energy gap $\mathcal{E}_{g2} > \mathcal{E}_{g1}$. If the thickness, d, of the layer with smaller bandgap is of the order of a few nanometers (for example, $d < 20$ nm), the confined electrons and holes show relevant quantum effects. The band structure of the material is greatly modified, thus leading to significant changes of all physical properties with respect to the bulk case, with very important advantages for various applications.

Another important difference with respect to the bulk case is the following. In the case of bulk semiconductors it is possible to have an epitaxial growth of one semiconductor onto another one only if the lattice spacings are almost the same for the two semiconductors, with maximum relative differences of the order of 0.1% of the lattice constant (for example, the relative lattice period variation, $\Delta a/a$, is $< 0.12\%$ in the case of GaAs/AlGaAs). In this way it is possible to fabricate semiconductor heterostructures. A GaAs/AlGaAs heterostructure can be easily obtained since the lattice period of GaAs (564 pm) is almost equal to the lattice period of AlAs (566 pm). In the case of the quaternary alloy $In_{1-x}Ga_xAs_yP_{1-y}$, lattice matching with InP can be achieved only for a specific ratio y/x. Nearly perfect lattice matching is required to minimize defects at the interface between the two semiconductors. A lattice mismatch leads to mechanical stress of the semiconductor, which may result in the generation of dislocation defects at the interface, with detrimental consequences on the optical and electronic properties of the material. The situation is different in the case of quantum wells. Due to the very small thickness of the QW layer, the lattice constants of the two semiconductors can be significantly different. As a result, strain develops within the QW layer. As it will be discussed in this chapter, the strain changes the properties of the QW semiconductor, in particular it changes the effective masses and

the bandedges, with potential remarkable advantages for the applications. We will see that the quantum confinement of electrons and holes in QW semiconductors and, for strained QWs, the variation of effective masses, lead to remarkable differences of the optical properties of QWs with respect to bulk semiconductors.

6.2 Electronic states

We will consider a QW consisting of a direct bandgap material, so that the bottom of the conduction band is of $s-$type, while the valence band states are of $p-$type. We will first consider the conduction band. Let us assume a potential well with thickness L_z (z is the direction of the semiconductor growth). The electronic wavefunction must be spatially localized within the well so that it can be written as a linear combination of Bloch wavefunctions:

$$\psi(\mathbf{r}) = \sum_{\mathbf{k}} F(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{r}} u_{\mathbf{k}}(\mathbf{r}). \tag{6.1}$$

In the framework of the *envelope function approximation* we can assume that the periodic function $u_{\mathbf{k}}(\mathbf{r})$ does not strongly depend on $\mathbf{k}$ around the bandedge, so that it can be approximated by its value near the bandedge ($\mathbf{k} = 0$), i.e., $u_{\mathbf{k}}(\mathbf{r}) \approx u_0(\mathbf{r})$. With this approximation the wavefunction $\psi(\mathbf{r})$ can be written as:

$$\psi(\mathbf{r}) = \left(\sum_{\mathbf{k}} F(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{r}} \right) u_0(\mathbf{r}) \equiv F(\mathbf{r}) u_0(\mathbf{r}), \tag{6.2}$$

where $F(\mathbf{r})$ is the so-called envelope function, so that the approximate solution of the Schrödinger equation can be written as the product of the bandedge Bloch function and a slowly varying envelope function. We can assume that both these functions are normalised so that:

$$\langle F|F\rangle = \int_V F^* F \, d\tau = 1 \tag{6.3}$$

and

$$\langle u_0|u_0\rangle = \frac{1}{V_u} \int_{\text{unit cell}} u_0^* u_0 \, d\tau = 1 \tag{6.4}$$

where V is the volume of the crystal and V_u is the volume of a unit cell of the crystal. Moreover, the functions $u_{0,c}(\mathbf{r})$ and $u_{0,v}(\mathbf{r})$, for the conduction and valence band, respectively, are orthonormal:

$$\langle u_{0,v}|u_{0,c}\rangle = 0. \tag{6.5}$$

In the case of a QW extending in the $z-$direction, the wavefunction $\psi(\mathbf{r})$ can be written as:

$$\psi(\mathbf{r}) = F_c(z) \exp[i(k_x x + k_y y)] u_0(x, y) = F_c(z) e^{i\mathbf{K}\cdot\mathbf{r}} u_0(x, y), \tag{6.6}$$

where $\mathbf{K} = k_x \mathbf{u}_x + k_y \mathbf{u}_y$ and the envelope function $F_c(z)$ approximately satisfies the following one-dimensional Schrödinger equation:

$$\left[-\frac{\hbar^2}{2m_{c,j}} \frac{\partial^2}{\partial z^2} + V_{c,j}(z) + V_{K,j}(z) \right] F_c(z) = \mathcal{E}_z F_c(z), \tag{6.7}$$

where the subscript j indicates if we are considering the well region ($j = w$) or the barrier region ($j = b$); $V_{c,j}(z)$ is the confining potential and $V_{K,j}(z)$ is given by:

$$V_{K,j}(z) = \frac{\hbar^2}{2m_{c,j}}(k_x^2 + k_y^2) = \frac{\hbar^2 K^2}{2m_{c,j}}. \tag{6.8}$$

In the well $V_{c,w}(z) = 0$, outside the well $V_{c,b}(z) = V_0$, as shown in Fig. 6.1. We

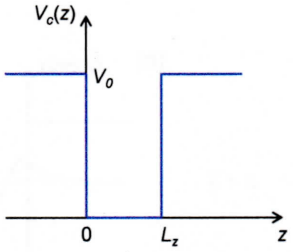

Figure 6.1: Schematic confining potential in a QW with thickness L_z.

will first consider the case of infinite barriers (i.e., $V_0 \to \infty$). As a first step we will assume that the electron has no momentum in the plane $x - y$, so that $V_{K,j} = 0$. In this case the boundary conditions are the following:

$$F_c(0) = F_c(L_z) = 0, \tag{6.9}$$

and Eq. 6.7 reduces to:

$$\frac{d^2F_c}{dz^2} = -\frac{2m_{c,w}}{\hbar^2}\mathcal{E}_z F_c = -k_z^2 F_c, \tag{6.10}$$

where:

$$k_z^2 = \frac{2m_{c,w}}{\hbar^2}\mathcal{E}_z. \tag{6.11}$$

The general solution is given by:

$$F_c(z) = A\sin(k_z z) + B\cos(k_z z). \tag{6.12}$$

Upon taking into account the boundary conditions 6.9 we have:

$$B = 0, \qquad A\sin(k_z L_z) = 0, \tag{6.13}$$

so that:

$$k_z = n\frac{\pi}{L_z}, \tag{6.14}$$

where n is an integer. From Eq. 6.11 we obtain:

$$\mathcal{E}_z = n^2\frac{\hbar^2\pi^2}{2m_{c,w}L_z^2} = \mathcal{E}_{nc}. \tag{6.15}$$

The coefficient A can be calculated by applying the normalization condition:

$$\int_0^{L_z} |F_c(z)|^2 dz = 1, \tag{6.16}$$

which gives $A = \sqrt{2/L_z}$, so that:

$$F_c(z) = \sqrt{\frac{2}{L_z}} \sin\left(n\frac{\pi}{L_z}z\right). \tag{6.17}$$

The first three functions, corresponding to $n = 1, 2$ and 3, are shown in Fig. 6.2(a). If the barriers of the potential well are not infinite, the wavefunction decays exponentially into the barrier regions. The first three solutions are shown in Fig. 6.2(b).

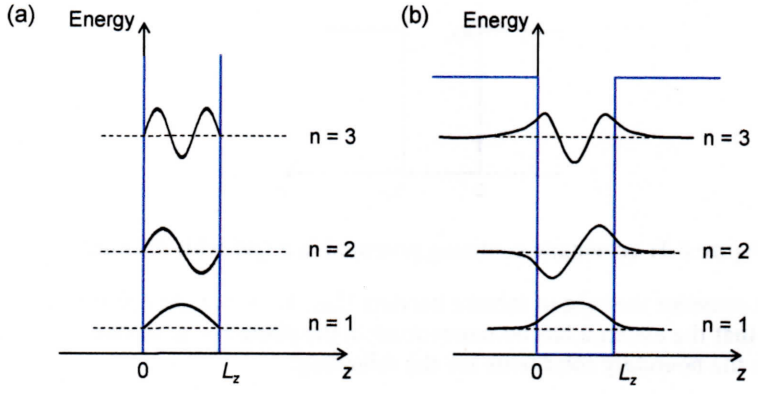

Figure 6.2: Eigenenergies and wavefunctions of the first three bound states of a quantum well in the case of infinite well depth (a) and of a well with finite potential barriers (b).

In Eq. 6.17 each n value corresponds to a two-dimensional subband in the conduction band, since the electron can move in the plane $x - y$ of the well. Indeed, the energy values given by Eq. 6.15 have been obtained upon assuming $V_{K,j} = 0$. In the general case, the total potential is given by $V_{c,j} + V_{K,j}$. The simplest case is when the effective masses in the well and in the barrier are equal: $m_w = m_b$. In this case $V_{K,j}$ does not depend on z and it just produces a change of the reference potential. If $m_w \neq m_b$ $V_{K,j}$ will depends on z and new solutions of Eq. 6.7 have to be calculated. In the limiting case when $V_{K,w} = V_0 + V_{K,b}$, the total potential profile $V_{c,j}(z) + V_{K,i}(z)$ is completely flat, thus removing any confinement effect. In general, we can assume that $|V_{K,w} - V_{K,b}| \ll V_0$, so that $V_{K,j}$ can be treated as a small perturbation of the potential well reported in Fig. 6.1. Therefore it is reasonable to assume that the solutions of Eq. 6.7 are close to the solutions found in the case $V_{K,j} = 0$, so that we can write $F_c(z) = F_{c,0}(z) + dF_c(z)$ and $\mathcal{E}_{nc} = \mathcal{E}_{nc,0} + d\mathcal{E}_{nc}$, where $F_{c,0}$ and $\mathcal{E}_{nc,0}$ are the unperturbed solutions of Eq. 6.7. These new solutions are then inserted in Eq. 6.7, which is then multiplied by $F^*_{c,0}$ and integrated from $-\infty$ to $+\infty$. After some calculations, the perturbation approach allows one to obtain the following approximated perturbed solutions:

$$\mathcal{E}_{nc} = \mathcal{E}_{nc,0} + \frac{\hbar^2}{2m_{eff}}(k_x^2 + k_y^2) = \mathcal{E}_{nc,0} + \frac{\hbar^2 K^2}{2m_{eff}}, \tag{6.18}$$

where the effective mass is given by:

$$\frac{1}{m_{eff}} = \frac{\Gamma}{m_w} + \frac{1-\Gamma}{m_b}, \tag{6.19}$$

and Γ is the confinement factor given by the following formula:

$$\Gamma = \int_0^{L_z} |F_{c,0}|^2 dz / \int_{-\infty}^{\infty} |F_{c,0}|^2 dz \tag{6.20}$$

The dispersion relation given by Eq. 6.18 is parabolic with an in-plane effective mass m_{eff}, which is a weighted average of the bulk effective masses in the well and in the barrier. Each level n gives rise to a two-dimensional subband in the $k_x - k_y$ plane, as shown in Fig. 6.3. The electron wavefunction can be written as:

$$\psi_c(x, y, z) = \sqrt{\frac{2}{L_z}} \sin\left(n\frac{\pi}{L_z}z\right) \exp(i\mathbf{K} \cdot \mathbf{r}) u_{c0}(\mathbf{r}), \tag{6.21}$$

where $\mathbf{K} \cdot \mathbf{r} = k_x x + k_y y$.

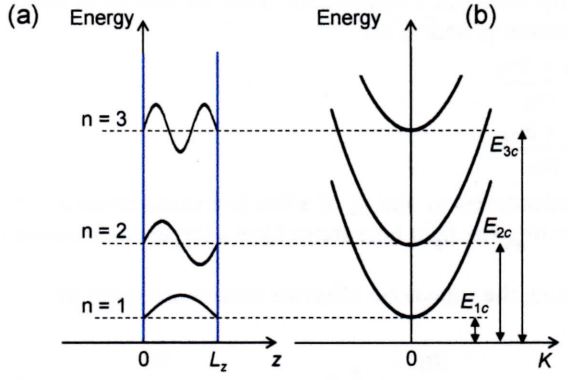

Figure 6.3: (a) Plots of the first three energy levels and of the corresponding eigenfunctions in the conduction band for infinite well depths (as in Fig. 6.2(a)); (b) corresponding energy vs K relations.

6.2.1 Electronic states in the valence band

In the case of the valence band the determination of the hole states when quantum confinement is present is more complicated. In the case of the electron states in the conduction band we have restricted our analysis to the case of direct band gap semiconductors, where the electronic states, at least near the bottom of the band, are described by a simple $s-$type band. In this case we can assume that the interaction with other energy bands is weak. As we have discussed in Ch. 1, the valence band states are generated from $p-$type states, thus leading to the formation of the heavy hole (HH) and light-hole (LH) bands, which are degenerate near the bandedge and whose reciprocal interaction cannot be generally treated as a small perturbation. By using the simple model employed in the case of the conduction band and assuming that the valence band may be represented by parabolic bands, the energy states in the well can be written as:

$$\mathcal{E}_{nv,hh} = \mathcal{E}_v - \mathcal{E}_{n,hh} - \frac{\hbar^2(k_x^2 + k_y^2)}{2m_{hh}}, \tag{6.22}$$

$$\mathcal{E}_{nv,lh} = \mathcal{E}_v - \mathcal{E}_{n,lh} - \frac{\hbar^2(k_x^2 + k_y^2)}{2m_{lh}}, \tag{6.23}$$

where $\mathcal{E}_v$ is the top of the valence band for the well semiconductor and the confinement energies $\mathcal{E}_{n,hh}$ and $\mathcal{E}_{n,lh}$ are given by:

$$\mathcal{E}_{n,hh} = n^2 \frac{\hbar^2 \pi^2}{2 m_{hh} L_z^2}, \tag{6.24}$$

$$\mathcal{E}_{n,lh} = n^2 \frac{\hbar^2 \pi^2}{2 m_{lh} L_z^2}. \tag{6.25}$$

Equations 6.22 and 6.23 show that in a QW the degeneracy between HH and LH states at the zone center is removed, since their effective mass enters in the confinement energy. The actual situation is much more complex as a result of the strong interaction between the valence bands. Here we just mention that a good approximation can be obtained by solving the Kohn-Luttinger form of the Schrödinger equation. Within this approach, the effective hole masses along the z direction (which have to be used in Eq. 6.7 with the subscript c replaced by v) can be written in terms of the so-called Luttinger parameters γ_1 and γ_2 as:

$$\frac{1}{m_{hh}^z} = \frac{\gamma_1 - 2\gamma_2}{m_0} \tag{6.26}$$

$$\frac{1}{m_{lh}^z} = \frac{\gamma_1 + 2\gamma_2}{m_0}. \tag{6.27}$$

The Luttinger parameters γ_1 and γ_2 of a few semiconductors are reported in Table 6.1. Since $m_{lh}^z < m_{hh}^z$ the light hole states have a larger confinement energy (see Fig. 6.4(a)).

In the $x - y$ plane, the transverse effective masses are given by:

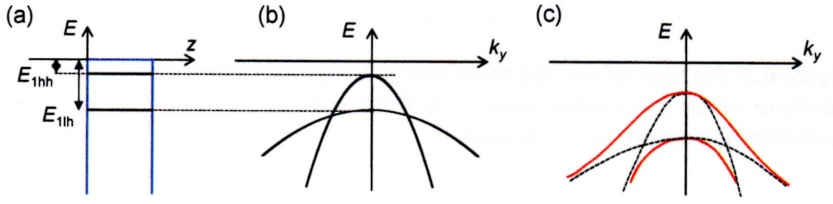

Figure 6.4: (a) The degeneracy between heavy and light hole states at $K = 0$ is removed in a QW. (b) Corresponding dispersion curves with band crossing; (c) high-order terms in the perturbative calculations remove the band crossing in a real QW (adapted from C. Weisbuch, *Fundamental Properties of III–V Semiconductor Two-Dimensional Quantized Structures: The Basis for Optical and Electronic Device Applications*, R. Dingle, ed., Semicond. Semimet. 24, 1 (1987))

$$\frac{1}{m_{hh}^{xy}} = \frac{\gamma_1 + \gamma_2}{m_0} \tag{6.28}$$

$$\frac{1}{m_{lh}^{xy}} = \frac{\gamma_1 - \gamma_2}{m_0}. \tag{6.29}$$

Note that in the $x - y$ plane, the mass of the heavy hole is smaller than the mass of the light holes, thus leading to the crossing points between the HH and LH bands shown in Fig. 6.4(b). The behavior in real QWs is more complex: indeed, higher order terms in the perturbation calculations remove such band crossing. Therefore the valence bands display an anticrossing behavior and can be strongly deformed, as schematically shown in Fig. 6.4(c).

	γ_1	γ_2
GaAs	6.98	2.06
AlAs	3.76	0.82
InAs	20.0	8.5
InP	5.08	1.60
GaN	2.67	0.75
AlN	1.92	0.47

Table 6.1: Luttinger parameters γ_1 and γ_2 for a few semiconductors. From I. Vurgaftman *et al.*, J. Appl. Phys. **89**, 5815-5875 (2001).

6.3 Density of states

In Sect. 2.2.2 the density of states has been calculated in the case of a two-dimensional structure:

$$\rho(\mathcal{E})d\mathcal{E} = \frac{K}{\pi}dK \tag{6.30}$$

where $\mathbf{K} = k_x\mathbf{u}_x + k_y\mathbf{u}_y$. For the conduction subband $n = 1$, assuming a parabolic dispersion curve:

$$\mathcal{E} = \mathcal{E}_{1c} + \frac{\hbar^2 K^2}{2m_{c,w}} \qquad d\mathcal{E} = \frac{\hbar^2 K}{m_{c,w}}dK \tag{6.31}$$

from Eq. 6.30 it is possible to obtain the density of states:

$$\rho_c(\mathcal{E}) = \frac{m_{c,w}}{\pi\hbar^2}. \tag{6.32}$$

The density of states can be written as a function of frequency, ω:

$$\rho_c(\mathcal{E}) = \frac{m_{c,w}}{\pi\hbar}. \tag{6.33}$$

The total density of states is given by the sum of the density of states from each subband:

$$\rho(\mathcal{E}) = \sum_n \frac{m_{c,w,n}}{\pi\hbar^2} H(\mathcal{E} - \mathcal{E}_{nc}), \tag{6.34}$$

where $H(\mathcal{E})$ is the Heaviside step function, as shown in Fig. 6.5.

It is interesting to note, as shown in Fig. 6.5, that the increase of the step-like function $\rho^{QW}(\mathcal{E})/L_z$ follows the function $\rho^{bulk}(\mathcal{E})$, i.e., the density of states of the bulk semiconductor, given by:

$$\rho^{bulk}(\mathcal{E}) = \frac{1}{2\pi^2}\left(\frac{2m_{c,w}}{\hbar^2}\right)^{3/2}\sqrt{\mathcal{E}}, \tag{6.35}$$

where we have assumed as a reference for the energy axis the bottom of the conduction band in the bulk case. Indeed:

$$\rho^{bulk}(\mathcal{E}_{nc}) = \rho^{bulk}(n^2\mathcal{E}_{1c}) = n\frac{m_{c,w}}{\pi\hbar^2 L_z} = \frac{\rho^{QW}}{L_z}. \tag{6.36}$$

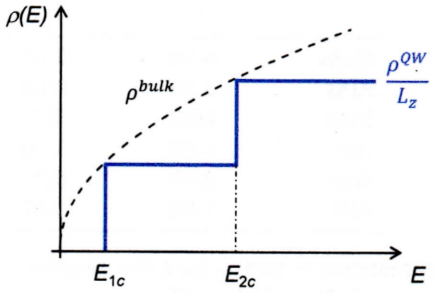

Figure 6.5: Plot of the quantum well density of states in the conduction band normalized to the well thickness, L_z, as a function of energy $\mathcal{E}_c$ (solid blue line) and plot of the density of states for the corresponding bulk semiconductor, ρ_{bulk} (dashed black line).

6.4 Electron density

At thermodynamic equilibrium, the probability of occupation of a particular state is given by the Fermi-Dirac distribution function. When electrons are excited in the conduction band it is possible to assume that the energy relaxation between the conduction subbands and between the valence subbands is very fast (in the picosecond or sub-picosecond time domain). As we have already seen, to calculate the electron and hole densities it is still possible to make use of the Fermi-Dirac distribution function by substituting the Fermi energy with the quasi-Fermi energies in the conduction and valence band:

$$n = \int \frac{\rho_c(\mathcal{E})}{L_z} f_c(\mathcal{E})\, d\mathcal{E} \tag{6.37}$$

$$p = \int \frac{\rho_v(\mathcal{E})}{L_z} \bar{f}_v(\mathcal{E})\, d\mathcal{E} \tag{6.38}$$

Since ρ is constant within each subband, the previous integrals can be easily calculated. Let us consider the $n-$subband in the conduction band:

$$\begin{aligned} n_n &= \frac{m_{c,w}}{\pi\hbar^2 L_z} \int_{\mathcal{E}_{nc}}^{\infty} \frac{d\mathcal{E}}{1+\exp\left(\frac{\mathcal{E}-\mathcal{E}_{Fc}}{k_B T}\right)} = \\ &= \frac{m_{c,w} k_B T}{\pi\hbar^2 L_z} \log\left[1+\exp\left(\frac{\mathcal{E}_{Fc}-\mathcal{E}_n}{k_B T}\right)\right], \end{aligned} \tag{6.39}$$

and the total electron density is given by:

$$n = \sum_n n_n \tag{6.40}$$

Exercise 6.1 Demonstrate that the density of electrons in the $n-$th subband of a quantum well is given by Eq. 6.39.

The integral in Eq. 6.39 can be easily calculated by using the variable x de-

fined as $x = \exp[(\mathcal{E} - \mathcal{E}_{Fc})/k_BT]$, so that:

$$dx = \frac{1}{k_BT}\exp\frac{\mathcal{E}-\mathcal{E}_{Fc}}{k_BT}\,d\mathcal{E} = \frac{x}{k_BT}\,d\mathcal{E}$$

$$n = \frac{m_{c,w}k_BT}{\pi\hbar^2 L_z}\int_{x_n}^{\infty}\frac{dx}{x(1+x)}$$

where $x_n = \exp[(\mathcal{E}_{nc} - \mathcal{E}_{Fc})/k_BT]$. Therefore:

$$\begin{aligned} n &= \frac{m_{c,w}k_BT}{\pi\hbar^2 L_z}\int_{x_n}^{\infty}\left(\frac{1}{x}-\frac{1}{1+x}\right)dx = \frac{m_{c,w}k_BT}{\pi\hbar^2 L_z}\log\left(1+\frac{1}{x_n}\right) \\ &= \frac{m_{c,w}k_BT}{\pi\hbar^2 L_z}\log\left[1+\exp\left(\frac{\mathcal{E}_{Fc}-\mathcal{E}_{nc}}{k_BT}\right)\right]. \end{aligned}$$

6.5 Transition selection rules

We will consider interband transitions from the valence band to the conduction band in a quantum well. The Fermi's Golden Rule is:

$$W_{12} = W_{vc} = \frac{2\pi}{\hbar}|H'_{cv}|^2\,\delta(\mathcal{E}_2 - \mathcal{E}_1 - \hbar\omega), \tag{6.41}$$

where the matrix element $|H'_{cv}|$ of the interaction Hamiltonian is given by:

$$H'_{cv} = \langle\psi_c|H'(\mathbf{r})|\psi_v\rangle = \int_V \psi_c^* H'(\mathbf{r})\psi_v\,d\tau, \tag{6.42}$$

and the interaction Hamiltonian is:

$$H'(\mathbf{r}) = -\frac{q}{m_0}\mathbf{A}\cdot\mathbf{p} = \frac{e}{m_0}\mathbf{A}(\mathbf{r})\cdot\mathbf{p}. \tag{6.43}$$

Considering an harmonic perturbation, the vector potential can be written as:

$$\mathbf{A}(\mathbf{r},t) = \hat{a}\frac{1}{2}\left[A(\mathbf{r})e^{-i\omega t} + A^*(\mathbf{r})e^{i\omega t}\right]. \tag{6.44}$$

where $A(\mathbf{r}) = \bar{A}(\mathbf{r})e^{i\mathbf{k}_{opt}\cdot\mathbf{r}}$. Due to the term $\delta(\mathcal{E}_2 - \mathcal{E}_1 - \hbar\omega)$ in Eq. 6.41 in the transition the energy is conserved:

$$\mathcal{E}_1 + \hbar\omega = \mathcal{E}_2. \tag{6.45}$$

The wavefunctions $\psi_{v,c}$ in a QW can be written as in Eq. 6.2, within the envelope function approximation, so that:

$$\begin{aligned} H'_{cv} &= \langle\psi_c|H'|\psi_v\rangle = \frac{e}{2m_0}\int_V F_c^* u_{c0}^*(A(\mathbf{r})\hat{a}\cdot\mathbf{p})F_v u_{v0}\,d\tau = \\ &= \frac{e}{2m_0}\int_V F_c^* u_{co}^*(A(\mathbf{r})\hat{a})u_{v0}\mathbf{p}F_v\,d\tau + \frac{e}{2m_0}\int_V F_c^* u_{c0}^*(A(\mathbf{r})\hat{a})F_v\mathbf{p}u_{v0}\,d\tau \end{aligned} \tag{6.46}$$

where we have used the following property ($\mathbf{p} = -i\hbar\nabla$):

$$\mathbf{p}(F_v u_{v0}) = u_{v0}\mathbf{p}F_v + F_v\mathbf{p}u_{v0}. \tag{6.47}$$

Equation 6.46 can be written as:

$$H'_{cv} = \frac{e}{2m_0}\int u^*_{c0}u_{v0}F^*_c[A(\mathbf{r})\hat{a}\cdot\mathbf{p}]F_v\,d\tau + \tag{6.48}$$
$$+ \frac{e}{2m_0}\int [F^*_c A(\mathbf{r})F_v]u^*_{c0}\,\mathbf{p}\cdot\hat{a}\,u_{v0}\,d\tau$$

Starting from the first integral, we can first calculate the integral over a single unit cell of the crystal and then sum the result considering all the unit cells in the crystal volume. It is reasonable to assume that the term $F^*_c[A(\mathbf{r})\hat{a}\cdot\mathbf{p}]F_v$ is approximately constant in each unit cell. In this case the first integral in Eq. 6.48 can be written as:

$$\frac{e}{2m_0}\sum_j [F^*_c(A(\mathbf{r}_j)\hat{a}\cdot\mathbf{p})F_v]\int_{\text{u.c.}} u^*_{c0}u_{v0}\,d\tau, \tag{6.49}$$

where the integral is calculated over a single unit cell (u.c.), j is an index which refers to a given unit cell and $\mathbf{r}_j$ is the position vector of the $j-$th cell. Since $\langle u_{0,v}|u_{0,c}\rangle = 0$ (see Eq. 6.5), the integral calculated over the unit cell is zero, so that the first integral in Eq. 6.48 is zero.
We can treat the second integral in Eq. 6.48 in the same way, so that it can be rewritten as:

$$H'_{cv} = \frac{e}{2m_0}\sum_j [F^*_c A(\mathbf{r}_j)F_v]\int_{\text{u.c.}} u^*_{c0}\,\hat{a}\cdot\mathbf{p}\,u_{v0}\,d\tau. \tag{6.50}$$

Since the integral over the unit cell is the same for each cell, it can be extracted from the sum:

$$H'_{cv} = \frac{e}{2m_0}\left(\frac{1}{V_u}\int_{\text{u.c.}} u^*_{c0}\,\hat{a}\cdot\mathbf{p}\,u_{v0}\,d\tau\right)\sum_j [F^*_c A(\mathbf{r}_j)F_v]\,V_u. \tag{6.51}$$

Assuming that the volume of the unit cell is very small, the sum in the previous expression can be written as an integral, so that:

$$H'_{cv} = \frac{e}{2m_0}\langle u_{c0}|\hat{a}\cdot\mathbf{p}|u_{v0}\rangle\int F^*_c A(\mathbf{r})F_v\,d\tau. \tag{6.52}$$

The envelope function overlap integral can be written as follows:

$$\int_V F^*_c A(\mathbf{r})F_v\,d\tau = \int_V F^*_c(z)e^{-i\mathbf{K}_c\cdot\mathbf{r}}\bar{A}(\mathbf{r})e^{i\mathbf{k}_{opt}\cdot\mathbf{r}}F_v(z)e^{i\mathbf{K}_v\cdot\mathbf{r}}\,d\tau, \tag{6.53}$$

this integral does not vanish if:

$$\mathbf{K}_c = \mathbf{K}_v + \mathbf{k}_{opt}. \tag{6.54}$$

Since $|\mathbf{k}_{opt}|$ is typically much smaller than $|\mathbf{K}_{v,c}|$, we can write the momentum selection rule 6.54 as:

$$\mathbf{K}_v = \mathbf{K}_c. \tag{6.55}$$

Since the well thickness is much smaller than the wavelength of the electromagnetic radiation, the vector potential can be assumed constant in the well region, so that the envelope function overlap integral can be written as:

$$\int_V F^*_c\bar{A}(\mathbf{r})F_v\,d\tau \simeq A_0\int F^*_c(z)F_v(z)\,dz = A_0\langle F_c|F_v\rangle. \tag{6.56}$$

Therefore the squared modulus of the matrix element H'_{cv} is given by:

$$|H'_{cv}|^2 = \left(\frac{eA_0}{2m_0}\right)^2 |\langle u_{c0}|\hat{a}\cdot\mathbf{p}|u_{v0}\rangle|^2 |\langle F_c|F_v\rangle|^2 = \left(\frac{eA_0}{2m_0}\right)^2 |M_T|^2 |\langle F_c|F_v\rangle|^2. \tag{6.57}$$

The term $|\langle F_v|F_c\rangle|^2$ leads to another transition rule. In the case of a well with infinite potential barriers we have calculated the functions $F_v(z)$ and $F_c(z)$:

$$F_{v,c} = \sqrt{\frac{2}{L_z}} \sin\left(n_{v,c}\frac{\pi}{L_z}z\right), \tag{6.58}$$

which are independent of the effective mass of electrons and holes. When $n_c = n_v$ the envelope functions $F(z)$ in the conduction and valence band are equal, so that $|\langle F_c|F_v\rangle|^2 = 1$: these transitions are allowed. On the contrary, when $n_c \neq n_v$, the envelope functions in the conduction and valence band are orthogonal to each other so that $|\langle F_c|F_v\rangle|^2 = 0$: these transitions are therefore strictly forbidden.
In the case of QWs with finite barriers, the wavefunction penetrates in the barrier regions so that the overall wavefunction weakly depends on the electron and hole effective masses. In this case when $n_c = n_v$ we still have $|\langle F_c|F_v\rangle|^2 \simeq 0.95 - 1$ but now if $n_c \neq n_v$ the envelope functions are not perfectly orthogonal, so that $|\langle F_c|F_v\rangle|^2 \simeq 0-0.1$ and the transition is not strictly forbidden although the probability of such transition is typically very small. Note that, also in the case of finite barriers, the transitions between a state with even envelope function and a state with odd envelope function are still strictly forbidden since in this case $|\langle F_c|F_v\rangle|^2 = 0$. As a final remark we add that, far from the bandedge, band mixing in the valence band becomes rather strong and the envelope functions $F_v(z)$ are not purely even or odd functions, so that most transitions are weakly allowed.
Note that the first integral in Eq. 6.48 can be written in another way (in analogy to what has been done for the second integral):

$$\frac{e}{2m_0}\int u^*_{c0}u_{v0}F^*_c[A(\mathbf{r})\hat{a}\cdot\mathbf{p}]F_v\, d\tau = \frac{eA_0}{2m_0}\langle u_{c0}|u_{v0}\rangle\langle F_c|\hat{a}\cdot\mathbf{p}|F_v\rangle. \tag{6.59}$$

6.5.1 Intersubband transitions

In a QW we can have *intraband* (or *intersubband*) transitions, i.e., transitions between different subbands either within the conduction band or the valence band. These intraband transitions have important applications in many photonic devices, ranging from far-infrared detectors to quantum cascade lasers. Let us consider intraband transitions in the conduction band of a QW grown, as always assumed in this chapter, along the $z-$direction. Let us consider two subbands, $n = 1, 2$. The corresponding wavefunctions can be written as:

$$\psi_{1c} = F_1(\mathbf{r})u_{c0}(\mathbf{r}) = F_1(z)e^{i\mathbf{K}\cdot\mathbf{r}}u_{c0}(\mathbf{r}) \tag{6.60}$$

$$\psi_{2c} = F_2(\mathbf{r})u_{c0}(\mathbf{r}) = F_2(z)e^{i\mathbf{K}\cdot\mathbf{r}}u_{c0}(\mathbf{r}) \tag{6.61}$$

where the central cell functions $u_{c0}(\mathbf{r})$ are the same for different subbands (this is valid in particular for the conduction band). We can use Eqs. 6.59 and 6.52 to write the matrix element H'_{21}:

$$H'_{21} = \frac{eA_0}{2m_0}\langle u_{c0}|u_{c0}\rangle\langle F_2|\hat{a}\cdot\mathbf{p}|F_1\rangle + \frac{eA_0}{2m_0}\langle u_{c0}|\hat{a}\cdot\mathbf{p}|u_{c0}\rangle\langle F_2|F_1\rangle. \tag{6.62}$$

As we have pointed out before, in the case of interband transitions the first term vanishes ($\langle u_{v0}|u_{c0}\rangle = 0$). In Eq. 6.62 the first term corresponds to the intersubband transition in the conduction band, while the second term vanishes due to the term $\langle F_2|F_1\rangle$, since $F_1(z)$ and $F_2(z)$ have even and odd parity, respectively (see Fig. 6.2(b)). Let us now analyze the first term in Eq. 6.62, which depends on the matrix element $\langle F_2|\hat{a}\cdot\mathbf{p}|F_1\rangle$, where $\mathbf{p} = -i\hbar\nabla$:

$$\langle F_2|\hat{a}\cdot\mathbf{p}|F_1\rangle = -i\hbar\int F_2^* e^{-i\mathbf{K}\cdot\mathbf{r}}\hat{a}\cdot\nabla F_1(z)e^{i\mathbf{K}\cdot\mathbf{r}}\,d\tau \tag{6.63}$$

If the polarization of the electromagnetic field is in the $x-y$ plane, the matrix element in Eq. 6.63 is zero: i.e., for light polarized in the $x-y$ plane the inter-subband transition rate is zero. Also in this case in the valence band this condition does not strictly hold as a consequence of the strong mixing of the central cell function. On the contrary, if the electromagnetic field has a $z-$polarized component, the matrix element in Eq. 6.63 does not vanish. Indeed, if the radiation is $z-$polarized Eq. 6.63 can be written as:

$$\langle F_2|\mathbf{u}_z\cdot\mathbf{p}|F_1\rangle = -i\hbar\int F_2^*(z)\mathbf{u}_z\cdot\left(\frac{dF_1(z)}{dz}\mathbf{u}_z\right)dz. \tag{6.64}$$

Since $F_1(z)$ and $F_2(z)$ have even and odd parity, respectively, $F_2(z)$ and $dF_1(z)/dz$ both have odd parity, therefore the integral in Eq. 6.64 does not vanish.

6.6 Absorption and gain in a quantum well

We will first calculate the joint density of states $\rho_j^{QW}(\mathcal{E}_0)$, where $\rho_j^{QW}(\mathcal{E}_0)\,d\mathcal{E}_0$ gives the number of allowed transitions per unit area with transition energy in the range between $\mathcal{E}_0$ and $\mathcal{E}_0 + d\mathcal{E}_0$. We will first consider interband transitions from the valence to the conduction band. Due to the transition selection rules derived in the previous section, and considering that the spin does not change in a transition, the number of transitions, $\rho_j^{QW}(\mathcal{E}_0)\,d\mathcal{E}_0$, is equal to the number of states (either in the valence or in the conduction band), $\rho_j^{QW}(K)\,dK$, corresponding to the same energy range. Let us assume a transition from the subband $n = 1$ in the valence band to the subband $n = 1$ in the conduction band:

$$\mathcal{E}_1 = -\mathcal{E}_{1v} - \frac{\hbar^2K^2}{2m_v} \tag{6.65}$$

$$\mathcal{E}_2 = \mathcal{E}_g + \mathcal{E}_{1c} + \frac{\hbar^2K^2}{2m_c} \tag{6.66}$$

where we have assumed an energy axis with the zero at the top of the valence band. The transition energy $\mathcal{E}_0$ is given by:

$$\mathcal{E}_0 = \mathcal{E}_2 - \mathcal{E}_1 = \mathcal{E}_g + \mathcal{E}_{1c} + \mathcal{E}_{1v} + \frac{\hbar^2K^2}{2m_r} \tag{6.67}$$

where m_r is the reduced mass. From Eq. 6.67 we observe that the minimum (or threshold) transition energy is given by $\hbar\omega = \mathcal{E}_g + \mathcal{E}_{1c} + \mathcal{E}_{1v} > \mathcal{E}_g$, so that the absorption in a QW is blue-shifted with respect to the absorption in the case of a bulk semiconductor: the energy difference is given by the sum of the confinement

energies in the valence and the conduction band and of the energy gap. The density of states in a QW is given by $\rho^{QW}(K) = K/\pi$, so that:

$$\rho_j^{QW}(\mathcal{E}_0)d\mathcal{E}_0 = \rho^{QW}(K)dK, \tag{6.68}$$

where $d\mathcal{E}_0 = \hbar^2 K dK/m_r$, thus obtaining:

$$\rho_j^{QW}(\mathcal{E}_0) = \frac{m_r}{\pi\hbar^2}, \qquad \rho_j^{QW}(\omega) = \frac{m_r}{\pi\hbar}. \tag{6.69}$$

The absorption coefficient from the subband $n = 1$ in the valence band to the subband $n = 1$ in the conduction band can be written by using the expression of $\alpha(\omega)$ obtained for bulk semiconductors (see Eq. 5.30) by replacing $\rho_j(\omega)$ with $\rho_j^{QW}(\omega)/L_z$ and by using Eq. 5.44 to write the squared modulus of the matrix element of the dipole moment in terms of the matrix element of the interaction Hamiltonian ($|\mu_{21}|^2 = 4|H'_{21}|^2/E_0^2 = 4|H'_{21}|^2/\omega^2 A_0^2$):

$$\begin{aligned}\alpha(\omega) &= \frac{\omega\pi}{cn\epsilon_0\hbar}|\mu_{21}|^2\frac{\rho_j^{QW}(\omega)}{L_z}[f_v(\hbar\omega) - f_c(\hbar\omega)] = \\ &= \frac{\omega\pi}{cn\epsilon_0\hbar}\frac{4|H'_{21}|^2}{\omega^2 A_0^2}\frac{\rho_j^{QW}(\omega)}{L_z}[f_v(\hbar\omega) - f_c(\hbar\omega)].\end{aligned} \tag{6.70}$$

By using Eq. 6.57 we can write:

$$\alpha(\omega) = \frac{\pi e^2}{cn\epsilon_0 m_0^2\omega\hbar}|M_T|^2\frac{m_r}{\pi\hbar L_z}[f_v(\hbar\omega) - f_c(\hbar\omega)], \tag{6.71}$$

or, in a more compact form:

$$\alpha(\omega) = \alpha_0(\omega)[f_v(\hbar\omega) - f_c(\hbar\omega)]. \tag{6.72}$$

If n_v and n_c are the quantum numbers identifying a given subband in the valence and conduction band, respectively, the absorption coefficient related to a given subband transition can be written as $\alpha_{sub}(\omega, n_v, n_c)$. The total absorption coefficient at a particular frequency ω is obtained by summing over all allowed subband transitions:

$$\alpha(\omega) = \sum_{n_v}\sum_{n_c}\alpha_{sub}(\omega, n_v, n_c). \tag{6.73}$$

Figure 6.6 shows the calculated absorption coefficient in a 10-nm GaAs/$Al_{0.3}Ga_{0.7}As$ QW for in-plane polarized light.

The gain coefficient is directly obtained from Eq. 6.72:

$$g(\omega) = \alpha_0(\omega)[f_c(\hbar\omega) - f_v(\hbar\omega)] \tag{6.74}$$

for a given interband transition. In order to have a net gain it is required that:

$$f_c(\hbar\omega) - f_v(\hbar\omega) > 0, \tag{6.75}$$

which implies the following relation:

$$\hbar\omega = \mathcal{E}_2 - \mathcal{E}_1 < \mathcal{E}_{Fc} - \mathcal{E}_{Fv} \tag{6.76}$$

which is the Bernard-Duraffourg condition we have already obtained in the case of bulk semiconductors. On the other hand we also have:

$$\hbar\omega = \mathcal{E}_2 - \mathcal{E}_1 > \mathcal{E}_g + \mathcal{E}_{1c} + \mathcal{E}_{1v}, \tag{6.77}$$

thus, in the case of QWs the transparency condition is defined as follows:

$$\mathcal{E}_{Fc} - \mathcal{E}_{Fv} = \mathcal{E}_g + \mathcal{E}_{1c} + \mathcal{E}_{1v}. \tag{6.78}$$

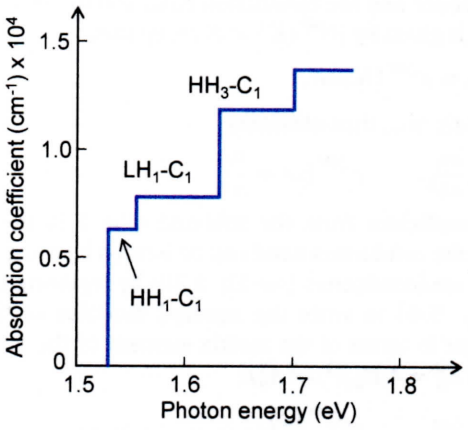

Figure 6.6: Calculated absorption coefficient in a 10-nm-thick GaAs/$Al_{0.3}Ga_{0.7}As$ QW for in-plane polarized light. Clearly visible the transitions from the heavy-hole (HH) subbands 1 and 3 to the conduction band ($n = 1$) and from the $n = 1$ light-hole subband to the $n = 1$ conduction subband. Note the presence of the HH_3-C_1 transition, which is not truly forbidden in the case of wells with finite potential barriers. Adapted from: J. Singh, Electronic and Optoelectronic Properties of Semiconductor Structures, Ch. 9, Cambridge University Press.

Exercise 6.2 Consider a GaAs well ($m_c = 0.067\ m_0$, $m_v = 0.45\ m_0$) with thickness $L_z = 10$ nm. Calculate the energy difference between the quasi-Fermi level in the conduction band and the bottom of the conduction band at transparency. Calculate the transparency density. Finally, assuming that $m_c = m_v = 0.067\ m_0$, calculate the corresponding transparency density.

At transparency $\mathcal{E}_{Fc} - \mathcal{E}_{Fv} = \mathcal{E}_g + \mathcal{E}_{1c} + \mathcal{E}_{1v}$, which can be written as $\mathcal{E}_{Fc} - \mathcal{E}_g - \mathcal{E}_{1c} = \mathcal{E}_{Fv} + \mathcal{E}_{1v} = \Delta\mathcal{E}$ (see Fig. 6.7). The electron density in the conduction band can be calculated by using Eq. 6.39:

$$n = \frac{m_c k_B T}{\pi\hbar^2 L_z} \log\left[1 + \exp\frac{\mathcal{E}_{Fc} - (\mathcal{E}_g + \mathcal{E}_{1c})}{k_B T}\right] = \\ = \frac{m_c k_B T}{\pi\hbar^2 L_z} \log\left(1 + \exp\frac{\Delta\mathcal{E}}{k_B T}\right). \tag{6.79}$$

If $\Delta\mathcal{E} > k_B T$ we can use the following approximated expression:

$$n \simeq \frac{m_c k_B T}{\pi\hbar^2 L_z} \frac{\Delta\mathcal{E}}{k_B T} = \frac{m_c}{\pi\hbar^2 L_z} \Delta\mathcal{E}. \tag{6.80}$$

The hole density in the valence band can be calculated by using the same procedure

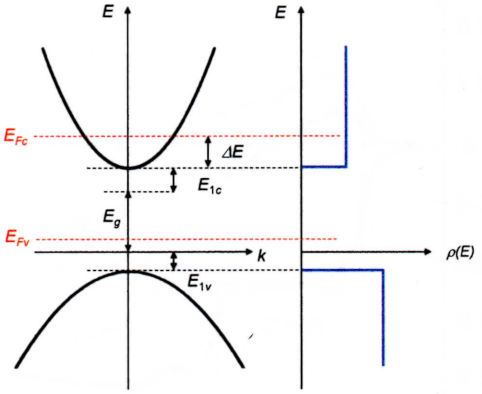

Figure 6.7: Dispersion curves of the $n = 1$ subbands in the valence and conduction band (solid black lines) and corresponding density of states (blue lines) at transparency.

used to obtain Eq. 6.39. We have:

$$p = \frac{m_v k_B T}{\pi\hbar^2 L_z} \log\left[1 + \exp\left(-\frac{\mathcal{E}_{1v} + \mathcal{E}_{Fv})}{k_B T}\right)\right] = \\ = \frac{m_v k_B T}{\pi\hbar^2 L_z} \log\left[1 + \exp\left(-\frac{\Delta\mathcal{E}}{k_B T}\right)\right]. \tag{6.81}$$

Again, if $\Delta\mathcal{E} > k_B T$ we can write:

$$p \simeq \frac{m_v k_B T}{\pi\hbar^2 L_z} \exp\left(-\frac{\Delta\mathcal{E}}{k_B T}\right). \tag{6.82}$$

Since, for charge neutrality we must have $n = p$, from Eqs. 6.80 and 6.82 we obtain:

$$\frac{\Delta\mathcal{E}}{k_B T} = \left(\frac{m_v}{m_c}\right) \exp\left(-\frac{\Delta\mathcal{E}}{k_B T}\right) = 6.7 \exp\left(-\frac{\Delta\mathcal{E}}{k_B T}\right), \tag{6.83}$$

which can be graphically solved. The result is $\Delta\mathcal{E} = 1.5\, k_B T = 39$ meV (at $T = 300$ K). Without approximations we would have obtained $\Delta\mathcal{E} = 1.33\, k_B T = 34.6$ meV (at $T = 300$ K). The corresponding electron density at transparency is $N_{tr} \simeq 1.1 \times 10^{18}$ cm^{-3}. Note that if $m_c = m_v$, $\Delta\mathcal{E} = 0$ and the transparency density is given by:

$$N_{tr} = \frac{m_c k_B T}{\pi\hbar^2 L_z} \log 2 \simeq 5 \times 10^{17} \text{ cm}^{-3}. \tag{6.84}$$

A typical evolution of the gain coefficient vs photon energy for different values of the injected carrier density is shown in Fig. 6.8, which refers to an unstrained GaAs/$Al_{0.2}Ga_{0.8}As$ QW (assuming parabolic subbands). The dashed curves have been obtained by considering a step-like joint density of states, as we have done so far, while the solid lines take into account the spectral broadening of each transition, due

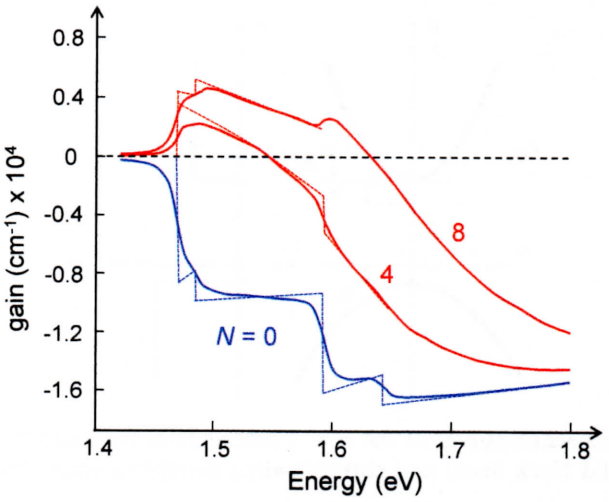

Figure 6.8: Gain coefficient as a function of the photon energy for different values of the injected carrier density, N (in units of 10^{18} cm^{-3}) for a 8-nm GaAs/$Al_{0.2}Ga_{0.8}As$ QW, in the parabolic band approximation. The dashed curves have been obtained by considering a step-like joint density of states, while the solid lines take into account the spectral broadening of each transition. The blue curve represents the absorption coefficient. Adapted from S.W. Corzine, R.H. Yan, and L.A. Coldren, *Optical Gain in III-V Bulk and Quantum well Semiconductors*, In: *Quantum Well Lasers* ed. by P.S. Zory (Academic, San Diego, CA, 1993) Chap. 1.

to scattering processes. The net effect of the lineshape broadening can be obtained by convolving a Lorentzian function with the calculated gain function obtained in this section. Better results can be achieved by replacing the Lorentzian function with another lineshape function obtained from a more exact calculation of the induced dipole phase damping.

Also in the case of QWs, above the transparency condition the peak of the gain increases with carrier injection. It is possible to write an approximated linear dependence of the gain peak vs carrier injection, as we have done for bulk semiconductors:

$$g_P = \sigma^{QW}(N - N_{tr}), \tag{6.85}$$

where σ^{QW} is the differential gain (of the order of $\sim 7 \times 10^{-16}$ cm^2 in the case of GaAs/$Al_{0.2}Ga_{0.8}As$ QWs). We note that in the case of QWs the linear approximation between g_P and N is not so good as in the case of bulk semiconductors.

6.7 Intersubband absorption

In the case of intersubband transitions, since the electronic subbands are parallel, all transitions between two subbands occur at the same energy (i.e., do not depend on **K**), therefore these interactions are resonant. The corresponding absorption coefficient can be calculated by using the same procedure used for bulk semiconductors. In the bulk case, the absorption coefficient associated to transitions with transition

frequency around ω_0 is given by Eq. 5.28:

$$d\alpha(\omega) = \frac{\omega\pi}{nc\epsilon_0\hbar}|\mu_{21}|^2 g(\omega-\omega_0)\{\rho_j(\omega_0)[f_v(\hbar\omega_0) - f_c(\hbar\omega_0)]d\omega_0\}. \tag{6.86}$$

In the case of intersubband transitions between two subbands in the conduction band the terms in the curly braces give the difference between the electronic density in the $n-$th (N_n) and the $m-$th (N_m) subbands, so that the absorption coefficient can be written as:

$$\alpha(\omega) = \frac{\omega\pi}{nc\epsilon_0\hbar}|\mu_{mn}|^2 g(\omega-\omega_{mn})(N_n - N_m), \tag{6.87}$$

where $\hbar\omega_{mn} = \mathcal{E}_m - \mathcal{E}_n$ and $g(\omega-\omega_{mn})$ is the Lorentzian lineshape of the transition. Here we have assumed an electromagnetic field with a $z-$ component. Let us consider a $n-$type semiconductor with $\mathcal{E}_n < \mathcal{E}_F < \mathcal{E}_m$ (where $\mathcal{E}_F$ is the Fermi energy). In this case we have $N_m \simeq 0$ and the absorption coefficient is given by:

$$\alpha(\omega) = \frac{\omega\pi}{nc\epsilon_0\hbar}|\mu_{mn}|^2 g(\omega-\omega_{mn})N_n, \tag{6.88}$$

which turns out to be proportional to the dopant concentration. If $\mathcal{E}_F > \mathcal{E}_m$ we can write (from Eq. 6.39):

$$\begin{aligned} N_n &= \frac{m_c k_B T}{\pi\hbar^2 L_z}\log\left[1+\exp\left(\frac{\mathcal{E}_F-\mathcal{E}_n}{k_B T}\right)\right] \simeq \frac{m_c}{\pi\hbar^2 L_z}(\mathcal{E}_F-\mathcal{E}_n) \\ N_m &\simeq \frac{m_c}{\pi\hbar^2 L_z}(\mathcal{E}_F-\mathcal{E}_m) \end{aligned} \tag{6.89}$$

and the absorption coefficient turns out to be independent of dopant concentration:

$$\alpha(\omega) = \frac{\omega\pi}{nc\epsilon_0\hbar}|\mu_{mn}|^2 g(\omega-\omega_{mn})\frac{m_c}{\pi\hbar^2 L_z}(\mathcal{E}_m-\mathcal{E}_n). \tag{6.90}$$

Exercise 6.3 Calculate the intersubband dipole moment in an infinite QW associated to the transition between the subband $n=1$ and $n=2$.

The envelope functions $F_{1,2}(z)$ are given by (see Eq. 6.17):

$$F_1(z) = \sqrt{\frac{2}{L_z}}\sin\left(\frac{\pi}{L_z}z\right), \tag{6.91}$$

$$F_2(z) = \sqrt{\frac{2}{L_z}}\sin\left(2\frac{\pi}{L_z}z\right). \tag{6.92}$$

The dipole matrix element μ_{21} can be calculated as follows:

$$\begin{aligned} \mu_{21} &= -e\int_0^{L_z} F_2(z) z F_1(z)\,dz = -\frac{2e}{L_z}\int_0^{L_z}\sin\left(\frac{\pi}{L_z}z\right) z \sin\left(2\frac{\pi}{L_z}z\right)\,dz = \\ &= -\frac{2eL_z}{\pi^2}\int_0^{\pi} x\sin(x)\sin(2x)\,dx = \\ &= -\frac{2eL_z}{\pi^2}\frac{1}{18}[12x\sin^3(x) + 9\cos(x) - \cos(3x)]_0^{\pi} = \frac{16}{9}\frac{eL_z}{\pi^2}. \end{aligned} \tag{6.93}$$

For example, assuming a well thickness $L_z = 10$ nm one obtains $\mu_{21} = 2.88 \times 10^{-28}$ C·m. The dipolar matrix element $|\langle F_2|z|F_1\rangle| = 1.8$ nm. ■

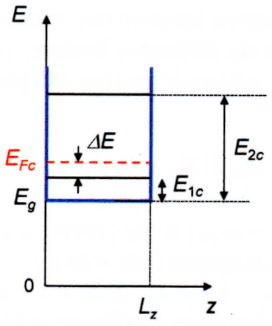

Figure 6.9: Schematic structure of the conduction band of the QW considered in Ex. 6.4.

Exercise 6.4 Calculate and plot the absorption spectrum corresponding to the intersubband transition from subband $n = 1$ to subband $n = 2$ in the conduction band of a GaAs QW with well thickness $L_z = 10$ nm for an injected electron density $N = 10^{18}$ cm^{-3}. Assume a dephasing time $T_2 = 110$ fs; the refractive index at the transition frequency is 3.3.

We first calculate the energy of the first two subbands in the conduction band (see Fig. 6.9):

$$\mathcal{E}_{1c} = \frac{\hbar^2\pi^2}{2m_c L_z^2} = 55.8 \text{ meV}$$
$$\mathcal{E}_{2c} = 4\mathcal{E}_{1c} = 223.1 \text{ meV}.$$

We can now calculate the position of the Fermi level in the conduction band. From Eq. 6.39 we have:

$$N = \frac{m_c k_B T}{\pi\hbar^2 L_z} \log\left[1 + \exp\left(\frac{\mathcal{E}_{Fc} - (\mathcal{E}_g + \mathcal{E}_{1c})}{k_B T}\right)\right] \tag{6.94}$$

The energy difference $\Delta\mathcal{E}$ between the Fermi level and the first state in the conduction band is given by $\Delta\mathcal{E} = \mathcal{E}_{Fc} - (\mathcal{E}_g + \mathcal{E}_{c1})$. From Eq. 6.94 we obtain:

$$\frac{\Delta\mathcal{E}}{k_B T} = \log\left[\exp\left(\frac{\pi\hbar^2 L_z}{m_c k_B T} N\right) - 1\right] = 1.08.$$

So that $\Delta\mathcal{E} = 1.08 k_B T = 28$ meV at $T = 300$ K. Therefore the Fermi energy is well below the subband $n = 2$ and the absorption coefficient is given by Eq. 6.88, with $N_n = N$:

$$\alpha(\omega) = \frac{\omega\pi}{nc\epsilon_0\hbar}|\mu_{21}|^2 g(\omega - \omega_{21})N, \tag{6.95}$$

where the matrix element μ_{21} is given by Eq. 6.93. Assuming a Lorentzian lineshape, the absorption peak is given by:

$$\alpha_P(\omega_0) = \frac{\omega_0 \pi}{nc\epsilon_0\hbar}\left(\frac{16}{9}\frac{eL_z}{\pi^2}\right)^2 \frac{T_2}{\pi} N \tag{6.96}$$

where T_2 is the dephasing time and ω_0 is the transition frequency $\omega_0 = (\mathcal{E}_{2c} - \mathcal{E}_{1c})/\hbar$. The wavelength corresponding to the resonant frequency is:

$$\lambda_0 = \frac{2\pi c}{\omega_0} = \frac{2\pi\hbar c}{\mathcal{E}_{c2} - \mathcal{E}_{c1}} = 7.4\ \mu\text{m}.$$

The absorption spectrum is shown in Fig. 6.10. Assuming a dephasing time $T_2 = 110$ fs the peak absorption is $\alpha_P \simeq 2.5 \times 10^4$ cm^{-1}. ■

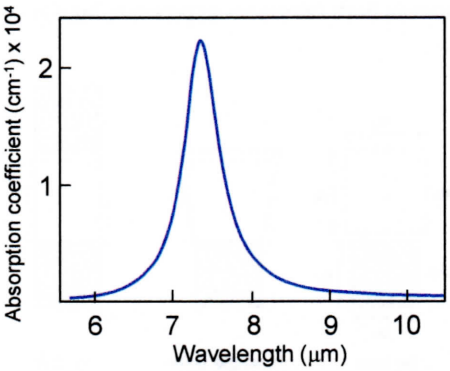

Figure 6.10: Absorption spectrum corresponding to the intersubband transition from subband $n = 1$ to subband $n = 2$ in the conduction band of the GaAs QW analyzed in Ex. 6.4.

6.8 Strained quantum wells

At the beginning of this chapter we have mentioned that in the case of bulk heterostructures obtained by growing a thick semiconductor layer above a different bulk semiconductor, an almost perfect lattice matching is required in order to prevent the formation of misfit dislocations. One of the most notable characteristics of the largely used GaAs/$Al_xGa_{1-x}As$ heterostructures is that the lattice periods of GaAs and AlAs are almost equal, with a relative lattice period variation $\Delta a/a < 0.12\%$. In this case different heterostructures can be used, with any combination of layer compositions without formation of dislocations. GaAs/AlGaAs heterostructures are used for the construction of double heterostructure lasers and QW heterostructure lasers with emission in the wavelength range between 650 nm and 800 nm, where the long wavelength limit is related to the bandgap of GaAs. Another largely used heterostructure is based on InP/$In_{1-x}Ga_xAs_yP_{1-y}$. The quaternary alloy $In_{1-x}Ga_xAs_yP_{1-y}$ can be lattice-matched to InP for a specific ratio $y/x \simeq 2.2$. Upon changing x, while keeping fixed the y/x ratio to the lattice-matching value, the semiconductor energy gap and therefore the emission wavelength can be changed between 1.15 μm and

1.67 μm for cw room temperature operation. A very good material for emission in the range between 880 nm and 1.1 μm is $In_xGa_{1-x}As$. In this case there are no suitable binary materials allowing lattice matched growth of $In_xGa_{1-x}As$. The unit cell of $In_xGa_{1-x}As$ can be as much as 3.6% larger (for $x = 0.5$) than that of GaAs. Therefore, considering a thick GaAs substrate, when $In_xGa_{1-x}As$ layers are deposited on top of GaAs the $In_xGa_{1-x}As$ cell is shortened in both directions parallel to the interface, thus giving a biaxial compression and it is elongated in the direction perpendicular to the interface, thus giving an uniaxial tension in this direction. Upon increasing the $In_xGa_{1-x}As$ layer thickness, the resulting strain energy increases, thus generating a force at the interface. When this force becomes larger than the tension in a dislocation line, a misfit dislocation is produced. In the case of $In_xGa_{1-x}As$ deposited on GaAs, for $x < 0.3$ the critical layer thickness, below which the strain produced by lattice mismatch does not lead to the formation of misfit dislocations, is greater than ~ 10 nm, so that it is possible to design QW devices properly working in a wide range of wavelengths.

In the QW configuration both interfaces experience biaxial compression, as shown in

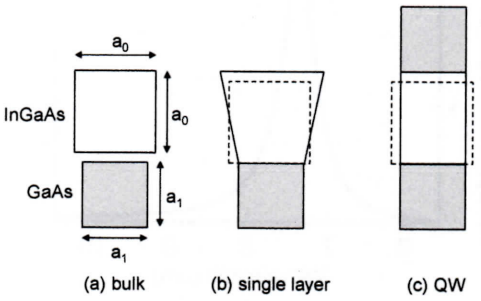

Figure 6.11: Crystal lattice deformation resulting from the epitaxial growth of a thin quantum-well layer of III–V material with original lattice constant a_0 (e.g., InGaAs) between two thick layers of a material with a lattice constant $a_1 < a_0$ (e.g., GaAs).

Fig. 6.11. The biaxial compression in the directions parallel to the interface destroys the cubic symmetry of the semiconductor. Since the energy gap of a semiconductor is related to its lattice spacing, we expect that a change of the bandgap of the strained layer is produced by the lattice distortion. The valence and conduction bands of strained semiconductor layers are modified by the presence of the strain, with a corresponding variation of the effective masses and the generation of energy shifts of the bandedges.

Soon after the development of the first strained QW heterostructures, many favourable characteristics of these structures were evidenced, ranging from lower threshold current density, higher differential gain and efficiency, increased range of available emission wavelengths. As a consequence, most of the commercially available semiconductor lasers are based on the use of strained semiconductors. Due to the remarkable practical importance of these structures, in this section we will analyze how strain affects the optical and electronic characteristics of III-V semiconductors.

The stress, or force per unit area, in a given crystal lattice is typically analysed in terms of a *stress tensor*, with nine components σ_{ij}, where the index i indicates that the stress is acting on a plane which is perpendicular to the x_i axis, and the index

j indicates the direction where the stress is acting. σ_{ii} are *normal stresses* and σ_{ij} are *shear stresses*. Figure 6.12 shows the components of the stress tensor in three

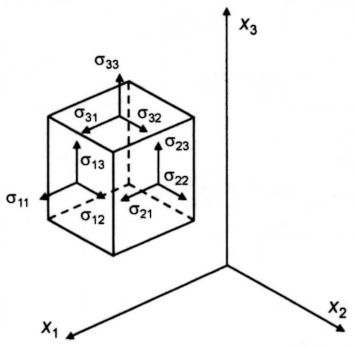

Figure 6.12: Components of the stress tensor in three dimensions.

dimensions. In semiconductor lattices employed in practical devices it is generally possible to assume that only normal stresses are present, i.e. $\sigma_{ij} = 0$ (for $i \neq j$). A stress component σ_{ii} is positive if it acts in the direction indicated by the outward unit vector parallel to the x_i axis. When a stress is applied to a lattice, strain is produced, which is also a second-rank tensor as the stress. Stress and strain are related by the following expression:

$$\sigma_{ij} = C_{ijkl}\, \epsilon_{kl}. \tag{6.97}$$

C_{ijkl} is called the elastic stiffness tensor and it is a fourth-rank tensor. Since it is generally sufficient to consider only normal stresses, we have to consider only the three diagonal components of the strain tensor, ϵ_{ii}, since the crystal axis, under the action of normal forces, remain orthogonal to each other. In the case of a cubic crystal the diagonal elements of the stiffness tensor are all equal and also the off-diagonal elements are all equal, and Eq. 6.97 can be written as:

$$\begin{bmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{zz} \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} & C_{12} \\ C_{12} & C_{11} & C_{12} \\ C_{12} & C_{12} & C_{11} \end{bmatrix} + \begin{bmatrix} \epsilon_{xx} \\ \epsilon_{yy} \\ \epsilon_{zz} \end{bmatrix} \tag{6.98}$$

so that we have only two elastic moduli, C_{11} and C_{12}. In typical semiconductors $C_{11}, C_{12} > 0$ and $C_{11} > C_{12}$. In the case of a biaxial strain (as in the case of InGaAs/GaAs heterostructures) the source stress is applied on the four faces perpendicular to the x and y axis, so that $\sigma_{xx} = \sigma_{yy}$, while no forces act on the two faces perpendicular to the z axis, so that $\sigma_{zz} = 0$. When the forces are directed outward (so that $\sigma_{xx} > 0$) we have a tensile strain (see Fig. 6.13(a)), while when the applied forces are directed inward ($\sigma_{xx} < 0$) we have a compressive strain (as shown in Fig. 6.13(b)). In Eq. 6.98, for symmetry reasons, since $\sigma_{xx} = \sigma_{yy}$, we also have $\epsilon_{xx} = \epsilon_{yy}$; $\sigma_{zz} = 0$ so that:

$$\sigma_{xx} = C_{11}\epsilon_{xx} + C_{12}\epsilon_{xx} + C_{12}\epsilon_{zz} \tag{6.99}$$

$$0 = C_{12}\epsilon_{xx} + C_{12}\epsilon_{xx} + C_{11}\epsilon_{zz} \tag{6.100}$$

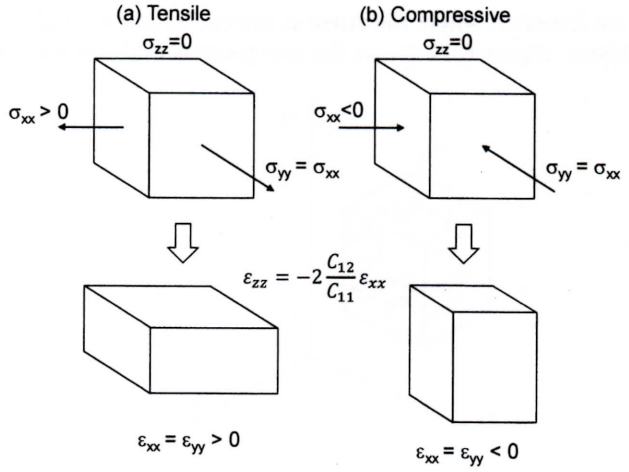

Figure 6.13: (a) Biaxial tensile strain and (b) biaxial compressive strain.

so that, from Eq. 6.100:

$$\epsilon_{zz} = -2\frac{C_{12}}{C_{11}}\epsilon_{xx}. \tag{6.101}$$

Therefore, since both C_{11} and C_{12} are positive, in the case of biaxial strain the lattice deformation along the z axis is opposite to the deformation along x and y axis. We recall that the sum of the diagonal elements of the strain tensor gives the volumetric strain, or fractional change of the volume of the crystal lattice (if we neglect the cross-terms between the ϵ_i)

$$\frac{dV}{V} \simeq \epsilon_{xx} + \epsilon_{yy} + \epsilon_{zz}. \tag{6.102}$$

In the following we will consider a biaxial strain. The strain $\epsilon = \epsilon_{xx} = \epsilon_{yy}$ is given by:

$$\epsilon = \frac{\Delta a}{a_0}, \tag{6.103}$$

where a_0 is the lattice constant of the substrate and Δa the variation of the lattice constant of the strained semiconductor. In the case of compressive strain $\Delta a < 0$ so that $\epsilon < 0$, while for tensile strain Δa and ϵ are both positive. It is possible to demonstrate that the variation of the bandgap induced by strain can be calculated by considering a suitable strain Hamiltonian. In particular, the bandgap variation can be expressed in terms of two energies, P and Q, which are related to the strain tensor through the deformation potentials a and b:

$$P = a(\epsilon_{xx} + \epsilon_{yy} + \epsilon_{zz}) \tag{6.104}$$

$$Q = b\left[\frac{1}{2}(\epsilon_{xx} + \epsilon_{yy}) - \epsilon_{zz}\right]. \tag{6.105}$$

By considering Eqs. 6.102 and 6.104 it is evident that P is proportional to the fractional change of the volume of the crystal lattice induced by the strain. This

energy term is called the hydrostatic component of the strain and a is also called hydrostatic deformation potential. In the case of biaxial strain:

$$P = 2a\epsilon\left(1 - \frac{C_{12}}{C_{11}}\right). \tag{6.106}$$

Q is called the shear component of the strain and gives the asymmetry in the strain parallel ($(\epsilon_{xx} + \epsilon_{yy})/2$) and perpendicular ($\epsilon_{zz}$) to the stress plane (do not confuse the shear strain with the shear stress, which is assumed to be zero in our discussion). b is also called tetragonal shear deformation potential. In the case of biaxial strain:

$$Q = b\epsilon\left(1 + 2\frac{C_{12}}{C_{11}}\right). \tag{6.107}$$

Assuming a biaxial strain in the $x - y$ plane the Hamiltonian can be simplified and the eigenvalues can be calculated. The variation in band gap is given by:

$$\Delta\mathcal{E}_g = P \tag{6.108}$$

while the splitting between the heavy and light hole bands is given by:

$$\Delta\mathcal{E}_{HH-LH} = 2Q. \tag{6.109}$$

Therefore, as shown schematically in Fig. 6.14, the variation of the energy difference

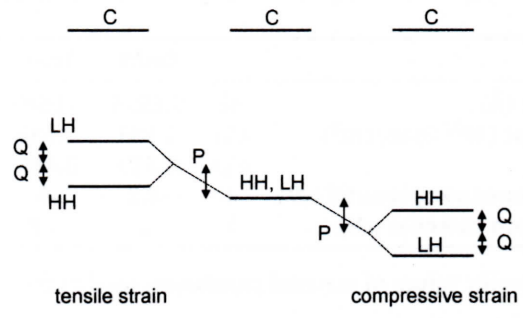

Figure 6.14: Schematic representation of the band energy shift for biaxial tensile and compressive strain. Although in this very schematic plot the conduction bandedge is not changed, this does not mean that the strain affects only the valence band. From the experimental point of view it is difficult to determine separately the shifts of the conduction and valence bandedges.

between the conduction band and valence band at $k = 0$ is given by:

$$\Delta\mathcal{E}_{HH} = P - Q = 2a\epsilon\left(1 - \frac{C_{12}}{C_{11}}\right) - b\epsilon\left(1 + 2\frac{C_{12}}{C_{11}}\right) \tag{6.110}$$

$$\Delta\mathcal{E}_{LH} = P + Q = 2a\epsilon\left(1 - \frac{C_{12}}{C_{11}}\right) + b\epsilon\left(1 + 2\frac{C_{12}}{C_{11}}\right) \tag{6.111}$$

where $\mathcal{E}_{HH}$ represents the shift of the heavy-hole valence bandedge with respect to the conduction bandedge and $\mathcal{E}_{LH}$ the corresponding shift of the light hole bandedge. In the case of biaxial compressive strain P and Q are both positive. For biaxial tensile

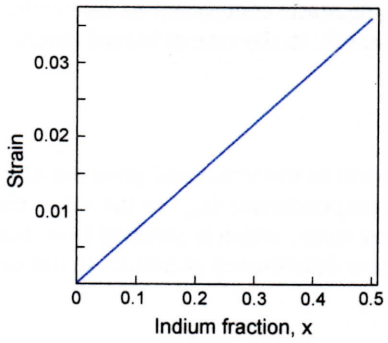

Figure 6.15: Variation of strain ϵ as a function of In fraction, x for an $In_xGa_{1-x}As/GaAs$ strained layer system.

strain P and Q are both negative.

It is instructive to analyse a particular example: $In_xGa_{1-x}As/GaAs$ (where the ternary alloy is the strained well semiconductor). The strain ϵ, given by Eq. 6.103, changes with In fraction x as shown in Fig. 6.15. The parameters used in Eqs. 6.110 and 6.111 for the calculation of the shifts in the energy of the bandedges can be interpolated from the corresponding values of the binary semiconductors GaAs ($x = 0$) and InAs ($x = 1$), which are given in Table 6.2. By using Eqs. 6.110 and

		GaAs	InAs	AlAs
Lattice constant (Å)	a_0	5.6535	6.0585	5.6600
Elastic coefficient (10^{12} dyne/cm^2)	C_{11}	1.188	0.865	1.25
	C_{12}	0.538	0.485	0.534
Hydrostatic deformation potential (eV)	a	-8.2	-6	-8.11
Shear deformation potential (eV)	b	-2	-1.8	-1.5

Table 6.2: List of specific values of material parameters used in the calculations.

6.111 and the numerical values, which can be obtained from Table 6.2, it is possible to calculate the increase of the energy separation between the conduction bandedge and the heavy-hole and light-hole valence bad edges at $k = 0$, which are reported in Fig. 6.16 as a function of x. Figure 6.16 shows that the shift of the light-hole valence band edge is much larger than the corresponding shift for the heavy-holes, therefore the optical transitions in strained layer $In_xGa_{1-x}As$ (e.g., recombination or absorption) are dominated by transitions between the conduction band and the heavy-hole valence band.

In a strained semiconductor, the strain does not just affect the positions of the band edges (in the conduction and valence bands) at $k = 0$. Figure 6.17 shows a schematic $\mathcal{E} - k$ diagram for the conduction and valence bands for an unstrained and for a compressively strained semiconductor layer, in the direction parallel and perpendicular to the growth interfaces. This figure clearly shows that the upper valence band (HH) is characterized by an effective mass greater than that corresponding to the lower valence band (LH) in the direction perpendicular to the plane of the interfaces, while it is characterized by a smaller effective mass

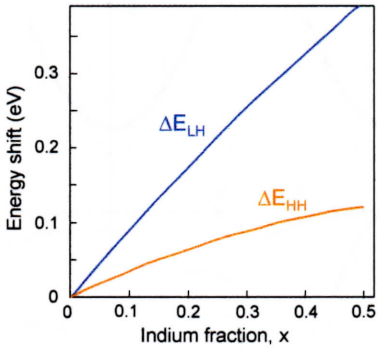

Figure 6.16: Shift of the heavy-hole valence bandedge, $\Delta\mathcal{E}_{HH}$, and of the light-hole valence band edge, $\Delta\mathcal{E}_{LH}$, with respect to the conduction band edge calculated by using Eqs. 6.110 and 6.111 vs indium fraction, x, for an $In_xGa_{1-x}As/GaAs$ strained layer system (adapted from J.J. Coleman, *Strained layer quantum well heterostructure lasers*, In: *Quantum Well Lasers* ed. by P.S. Zory (Academic, San Diego, CA, 1993) Chap. 8).

in the direction parallel to the interfaces. Note that the density of states depends on the effective masses in the direction parallel to the interfaces, which therefore directly affects the optical gain and the threshold current density in strained layer QW heterostructure lasers, as it will be explained in the next section.

6.9 Transparency density and differential gain

We have seen that in order to achieve a net gain the Bernard-Duraffourg condition must be satisfied. It is obvious that if the semiconductor is in thermal equilibrium $\mathcal{E}_{Fc} = \mathcal{E}_{Fv}$ and it is not possible to achieve a net gain. In the case of non-degenerate semiconductors $f_c(\mathcal{E}_2) \simeq 0$ and $f_v(\mathcal{E}_1) \simeq 1$, so that photons with energy $\hbar\omega = \mathcal{E}_2 - \mathcal{E}_1$ are strongly absorbed. In the case of a degenerate n^+-type semiconductor (still in thermal equilibrium) we have $f_c(\mathcal{E}_2) \simeq 1$ and $f_v(\mathcal{E}_1) \simeq 1$, so that the material is almost transparent for photons with energy $\hbar\omega = \mathcal{E}_2 - \mathcal{E}_1$. Therefore, we conclude that net gain requires a non-equilibrium condition in the semiconductor in order to produce simultaneously a high concentration of electrons in the conduction band and a high concentration of holes in the valence band. This condition can be achieved for example in the depletion region of a $p-n$ junction under forward bias or by optical pumping, which generates a high density of electron-hole pairs.

The minimum carrier density required to have a positive gain is the transparency density, N_{tr}, which is related to the density of states in the valence and conduction bands. In Ex. 6.2 the transparency carrier density has been calculated in the case of a GaAs QW with $m_v/m_c \simeq 7.5$: in this case the density of states in the valence and conduction bands are strongly asymmetric. We have seen that in this case the quasi-Fermi level $\mathcal{E}_{Fc}$ moves into the conduction band and $\mathcal{E}_{Fv}$ moves outside the valence band in order to preserve charge neutrality. In general, the effective mass asymmetry shifts both quasi-Fermi levels towards the band characterized by the lighter effective mass. The asymmetry leads to an increase of the transparency density. Indeed, as calculated in Ex. 6.2, in the case of a symmetric QW ($m_c = m_v$), the transparency density decreases by a factor of ~ 2.4. As we have discussed in Sec.

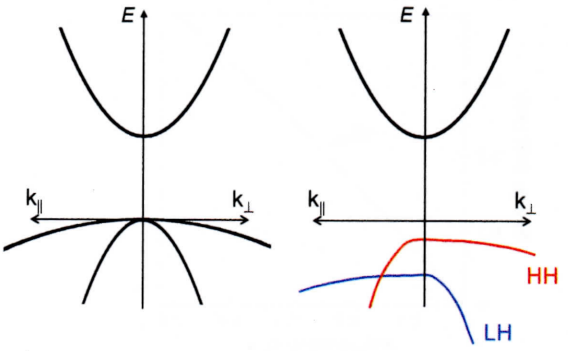

Figure 6.17: Schematic $\mathcal{E} - k$ diagram for the conduction and valence bands for an unstrained and for a compressively strained semiconductor layer, in the direction parallel and perpendicular to the growth interfaces (adapted from J.J. Coleman, IEEE J. Quantum Electr. **6**, 1008 (2000)).

6.8, the use of strained QWs allows one to significantly decrease the effective mass in the valence band, so that symmetric structures can be achieved. Note that, in terms of transparency density, the best situation is obtained in the case of a symmetric bulk semiconductor, since in this case the density of states smoothly increases from zero both in the valence and in the conduction band.

Exercise 6.5 Consider a symmetric bulk semiconductor with effective masses $m_c = m_v$ equal to the conduction effective mass of GaAs ($m_c = 0.067\ m_0$). Calculate the transparency carrier density.

Since the semiconductor is symmetric, the position of the quasi-Fermi levels at transparency coincides with the band edges. The density of electrons in the conduction band at transparency can be calculates as follows ($\mathcal{E}_{Fc} = \mathcal{E}_c$ and $\mathcal{E}_{Fv} = \mathcal{E}_v$):

$$N_{tr} = \int_{\mathcal{E}_c}^{\infty} \rho_c(\mathcal{E}) f(\mathcal{E})\, d\mathcal{E}$$

Therefore we have:

$$N_{tr} = \frac{1}{2\pi^2} \left(\frac{2m_c}{\hbar^2} \right)^{3/2} \int_{\mathcal{E}_c}^{\infty} \frac{\sqrt{\mathcal{E} - \mathcal{E}_c}}{1 + \exp(\mathcal{E} - \mathcal{E}_c)/k_B T}\, d\mathcal{E} \tag{6.112}$$

If, as usual, the parameter $x = (\mathcal{E} - \mathcal{E}_c)/k_B T$ is used in Eq. 6.112, the following result is obtained:

$$N_{tr} = N_c F_{1/2}(0) \tag{6.113}$$

where $F_{1/2}(y)$ is the Fermi function introduced in Sect. 2.3 and $F_{1/2}(0) = 0.765$. Since $N_c = 4.4 \times 10^{17}\ \text{cm}^{-3}$ (see Ex. 2.3), we obtain $N_{tr} = 3.37 \times 10^{17}\ \text{cm}^{-3}$, which is about 1.5 times smaller than the transparency density calculated in the

case of a symmetric QW with the same effective mass in the conduction and valence band.
■

We can conclude that, in order to minimize the transparency carrier density, it is useful to use symmetric semiconductors with small density of states and a "soft" band edge.

6.9.1 Differential gain

Another very important parameter is the differential gain, $\sigma = dg/dN$, which determines how the gain of a given semiconductor device reacts to a variation of the injected carrier density. The gain coefficient g is proportional to $\rho_j(\mathcal{E})[f_c(\mathcal{E}) - f_v(\mathcal{E})]$, which can be written as:

$$g \propto \rho_j(\mathcal{E})f_c(\mathcal{E}) + \rho_j(\mathcal{E})[1 - f_v(\mathcal{E})] - \rho_j(\mathcal{E}). \tag{6.114}$$

In symmetric structures the joint density of states is almost equal to the density of states (in the conduction or valence band) since the reduced mass m_r contained in the expression of $\rho_j(\mathcal{E})$ can be replaced by $m_c/2$ (or $m_v/2$). Therefore in Eq. 6.114, the first term on the right side, $\rho_j(\mathcal{E})f_c(\mathcal{E})$ is directly related to the electron density in the conduction band, while the second term, $\rho_j(\mathcal{E})[1 - f_v(\mathcal{E})]$, is directly related to the hole density in the valence band. A high differential gain corresponds to a high variation of the bandedge carrier density in response to a given increase in the separation between the quasi-Fermi levels, starting from the transparency condition. Figures 6.18(a) and 6.18(b) show the position of quasi-Fermi levels at transparency and the corresponding carrier density (displayed by the shaded areas) in the case of two symmetric structures (bulk and QW). It is evident that an increase of the

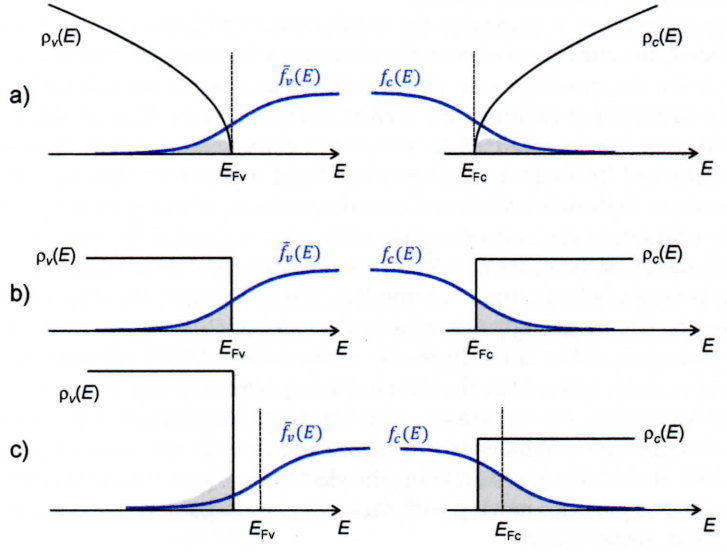

Figure 6.18: Calculation of the transparency carrier density (represented by the shaded areas) in the case of (a) a symmetric bulk semiconductor; (b) a symmetric QW and (c) an asymmetric QW.

separation between the two quasi-Fermi levels is more effective in the case of the symmetric QW (6.18(b)) due to the step-like character of the corresponding density of states. Therefore, it is reasonable to conclude that a symmetric QW is characterized by a higher differential gain in comparison to a symmetric bulk structure. On the other hand, it is also evident that the differential gain of a symmetric QW is higher than the differential gain of an asymmetric QW (shown in Fig. 6.18(c)). Indeed, in the symmetric case at transparency the maximum slope of the Fermi-Dirac distribution function coincides with the band edges, so that the band edge carrier density is very sensitive to a variation of the position of the quasi-Fermi level with respect to the transparency condition. The situation is completely different in the case of an asymmetric structure, since in this case the band edges do not correspond to the maximum slope of the Fermi-Dirac function, so that the band edge carrier density turns out to be much less sensitive to a variation of the position of the quasi-Fermi levels, thus resulting in a smaller differential gain. In conclusion, in terms of differential gain it is useful to use a symmetric band structure with step edges. Also in this case the use of strained QWs turns out to be beneficial to improve the performances of semiconductor devices.

6.10 Excitons

Before closing this chapter it is useful to briefly discuss important effects on the optical properties of a semiconductor related to electron-electron interactions. So far we have always adopted the single-electron approximation, which is very useful since it allowed us to investigate several electronic and optical properties of a semiconductor by using a single-particle Schrödinger equation. In the following we will consider only the Coulomb component of the electron-electron interaction and we will neglect both exchange and correlation terms.

When an electron is excited to the conduction band a hole is generated in the valence band; the attraction between the electron an the hole gives rise to a correlated motion of the electron-hole pair, which behaves as a single particle called *exciton*. There are two types of excitons: the *Frenkel excitons* and the *Wannier-Mott excitons*. The Frenkel exciton is localized to a few unit cells, therefore it is delocalized in the $k-$space and its analysis requires a full-band treatment. This type of exciton is important in materials with low dielectric constant, where screening effects are weak, for example in the case of organic molecules. A Frenkel exciton generally has a binding energy of the order of 0.1 to 1 eV. In semiconductors, where the dielectric constant is typically large, the Coulomb interaction between the electron and hole in the exciton can be strongly screened by the valence electrons, so that the exciton binding energy is rather small (typically of the order of 0.01 eV) and the exciton dimension is much larger than the lattice spacing, with a typical dimension of a few hundreds Angstroms. For this reason, near bandedge electron and hole states can be used to describe these Wannier-Mott excitons and the effective mass approximation can be used. Within this approximation, the electron and hole which form the exciton are described as particles moving with their effective masses, in the conduction and valence band, respectively.

The simplest approach for the investigation of excitonic effects in semiconductors is based on the analogy with a hydrogen atom, which we have already adopted in Sect. 2.5 for the description of the donor and acceptor states, which can be seen as particular "excitons", where one of the two coupled particles has an infinite

mass. In analogy with the hydrogen atom problem, we can assume, as a first order approximation, that the presence of excitons leads to the formation of discrete states within the gap of the semiconductor, with energies given by the following expression:

$$\mathcal{E} = \mathcal{E}_g - \mathcal{E}_{Bn} \tag{6.115}$$

where $\mathcal{E}_g$ is the energy gap and $\mathcal{E}_{Bn}$ is the binding energy of the $n-$th state of the exciton, given by:

$$\mathcal{E}_{Bn} = \frac{m_r e^4}{2(4\pi\epsilon\hbar)^2}\frac{1}{n^2}, \tag{6.116}$$

where we have used the reduced mass of the electron-hole pair and the dielectric constant of the semiconductor ($\epsilon = \epsilon_0\epsilon_r$). The exciton radius is given by:

$$R_B = \frac{4\pi\epsilon\hbar^2}{m_r e^2}. \tag{6.117}$$

By using Eqs. 6.116 and 6.117 we can calculate a first-order approximation for the energy of the excitonic state $n = 1$ in a semiconductor and the corresponding radius. In the case of GaAs: $m_c = 0.067\, m_0$, $m_{hh} = 0.45\, m_0$, $\epsilon_r = 12.9$ so that $m_r = 0.058\, m_0$, $\mathcal{E}_{B1} = 4.7$ meV and $R_B = 117$ Å. Note that the lattice constant of GaAs is $a_0 = 5.6$ Å, so that the exciton Bohr radius is more than 20 times larger, thus justifying the use of the effective mass approximation for the description of the Wannier exciton.

We will now study the exciton motion in terms of the motion of its centre of mass (CM) and of the relative motion of the electron and hole about the exciton CM. We anticipate that, within the effective mass approximation, the exciton CM moves as a free particle with mass $M = m_e + m_h$, where m_e and m_h are the electron and hole masses, respectively, while the relative motion of the electron and hole is similar to the relative motion of an electron and proton in an hydrogen atom. In the effective mass approximation we have to solve the following two-particle Schrödinger equation, which contains the electron-hole Coulomb interaction:

$$\left(-\frac{\hbar^2}{2m_e}\nabla^2_{\mathbf{r}_e} - \frac{\hbar^2}{2m_h}\nabla^2_{\mathbf{r}_h} - \frac{e^2}{4\pi\epsilon|\mathbf{r}_e - \mathbf{r}_h|}\right)\psi(\mathbf{r}_e, \mathbf{r}_h) = \mathcal{E}\psi(\mathbf{r}_e, \mathbf{r}_h) \tag{6.118}$$

In Eq. 6.118 $\mathbf{r}_e$ and $\mathbf{r}_h$ are the position vectors of the electron and hole, so that $|\mathbf{r}_e - \mathbf{r}_h|$ represents their distance. The effective mass equations for a free electron and a free hole are:

$$-\frac{\hbar^2}{2m_e}\nabla^2_{\mathbf{r}_e}\psi_c(\mathbf{r}_e) = \mathcal{E}_c\psi_c(\mathbf{r}_e) \tag{6.119}$$

$$-\frac{\hbar^2}{2m_h}\nabla^2_{\mathbf{r}_h}\psi_v(\mathbf{r}_h) = \mathcal{E}_v\psi_v(\mathbf{r}_h). \tag{6.120}$$

Without the Coulomb interaction term in Eq. 6.118 we would have: $\psi(\mathbf{r}_e, \mathbf{r}_h) = \psi_c(\mathbf{r}_e)\psi_v(\mathbf{r}_h)$. In the following we will always assume a weak Coulomb interaction due to the screening effect of the valence electrons, so that the effective mass approximation is valid. Moreover we will assume:

$$\mathcal{E}_c(\mathbf{k}_e) = \mathcal{E}_g + \frac{\hbar^2 k_e^2}{2m_e} \tag{6.121}$$

$$\mathcal{E}_v(\mathbf{k}_h) = -\frac{\hbar^2 k_h^2}{2m_h} \tag{6.122}$$

If we now consider the effects of the Coulomb interaction between the electron and hole, it is convenient to use the following transformations:

$$\mathbf{R} = \frac{m_e \mathbf{r}_e + m_h \mathbf{r}_h}{M} \tag{6.123}$$

which gives the position of the exciton CM, and:

$$\mathbf{r} = \mathbf{r}_e - \mathbf{r}_h \tag{6.124}$$

which gives the relative position of the electron with respect to the hole. Moreover we can introduce the center of mass wavevector, $\mathbf{K}$, as follows:

$$\mathbf{K} = \mathbf{k}_e + \mathbf{k}_h. \tag{6.125}$$

By using Eqs. 6.123 and 6.124 in the Schrödinger equation 6.118 we obtain:

$$\left(-\frac{\hbar^2}{2M}\nabla^2_{\mathbf{R}} - \frac{\hbar^2}{2m_r}\nabla^2_{\mathbf{r}} - \frac{e^2}{4\pi\epsilon r}\right)\psi(\mathbf{R},\mathbf{r}) = \mathcal{E}\psi(\mathbf{R},\mathbf{r}). \tag{6.126}$$

Exercise 6.6 Show that the Schrödinger equation 6.118 can be written as in Eq. 6.126.

Since $\mathbf{r} = \mathbf{r}_e - \mathbf{r}_h$, the components of $\mathbf{r}$ along x, y and z axis are: $x = x_e - x_h$, $y = y_e - y_h$ and$z = z_e - z_h$, while the components of $\mathbf{R}$ are:

$$X = \frac{m_e x_e + m_h x_h}{M}$$
$$Y = \frac{m_e y_e + m_h y_h}{M}$$
$$Z = \frac{m_e z_e + m_h z_h}{M}$$

The derivative with respect to x_e can be written as:

$$\frac{\partial}{\partial x_e} = \frac{\partial}{\partial x}\frac{\partial x}{\partial x_e} + \frac{\partial}{\partial X}\frac{\partial X}{\partial x_e} = \frac{\partial}{\partial x} + \frac{m_e}{M}\frac{\partial}{\partial X}$$

In the same way:

$$\begin{aligned}\frac{\partial^2}{\partial x_e^2} &= \frac{\partial}{\partial x_e}\left(\frac{\partial}{\partial x_e}\right) = \frac{\partial}{\partial x}\left(\frac{\partial}{\partial x_e}\right)\frac{\partial x}{\partial x_e} + \frac{\partial}{\partial X}\left(\frac{\partial}{\partial x_e}\right)\frac{\partial X}{\partial x_e} = \\ &= \frac{\partial}{\partial x}\left(\frac{\partial}{\partial x} + \frac{m_e}{M}\frac{\partial}{\partial X}\right) + \frac{\partial}{\partial X}\left(\frac{\partial}{\partial x} + \frac{m_e}{M}\frac{\partial}{\partial X}\right)\frac{m_e}{M} = \\ &= \frac{\partial^2}{\partial x^2} + \left(\frac{m_e}{M}\right)^2\frac{\partial^2}{\partial X^2} + 2\frac{m_e}{M}\frac{\partial^2}{\partial x \partial X}.\end{aligned}$$

In the same way we can write:

$$\frac{\partial^2}{\partial x_h^2} = \frac{\partial^2}{\partial x^2} + \left(\frac{m_h}{M}\right)^2\frac{\partial^2}{\partial X^2} - 2\frac{m_h}{M}\frac{\partial^2}{\partial x \partial X}$$

By repeating the same procedure for $\partial^2/\partial y_e^2$, $\partial^2/\partial y_h^2$, $\partial^2/\partial z_e^2$ and $\partial^2/\partial z_h^2$ we obtain:

$$-\frac{\hbar^2}{2m_c}\nabla^2_{\mathbf{r}_e} - \frac{\hbar^2}{2m_v}\nabla^2_{\mathbf{r}_h} = -\frac{\hbar^2}{2M}\nabla^2_{\mathbf{R}} - \frac{\hbar^2}{2m_r}\nabla^2_{\mathbf{r}},$$

so that the complete Schrödinger equation can be written as in Eq. 6.126. ■

In this way we have written an equation where the coordinates of the exciton CM and the relative coordinates are decoupled, so that we can write the wavefunction $\psi(\mathbf{R}, \mathbf{r})$ as a product of two wavefunctions:

$$\psi(\mathbf{R}, \mathbf{r}) = F(\mathbf{R})g(\mathbf{r}). \tag{6.127}$$

In this way from Eq. 6.126 we can write two equations:

$$-\frac{\hbar^2}{2M}\nabla^2_{\mathbf{R}}F(\mathbf{R}) = \mathcal{E}_R F(\mathbf{R}), \tag{6.128}$$

which describes the motion of the exciton CM and:

$$\left(-\frac{\hbar^2}{2m_r}\nabla^2_{\mathbf{r}} - \frac{e^2}{4\pi\epsilon r}\right) g(\mathbf{r}) = \mathcal{E}_r g(\mathbf{r}), \tag{6.129}$$

which describes the relative motion of the electron about the hole. The total energy of the exciton is $\mathcal{E}_R + \mathcal{E}_r$. We note that Eq. 6.128 is the Schrödinger equation of a free particle with mass M, so that:

$$F(\mathbf{R}) = \frac{1}{\sqrt{V}}\exp(i\mathbf{K}\cdot\mathbf{R}) \tag{6.130}$$

$$\mathcal{E}_R = \frac{\hbar^2 K^2}{2M} \tag{6.131}$$

this last equation represents the kinetic energy due to the motion of the exciton CM. Equation 6.129 is the same equation which describes the hydrogen atom, so that the wavefunction $g(\mathbf{r})$ can be written, in polar coordinate (r, θ, ϕ), as:

$$g_{n\ell m}(\mathbf{r}) = R_{n\ell}(r)Y_{\ell m}(\theta, \phi), \tag{6.132}$$

where n is the principal quantum number, ℓ is the angular momentum quantum number and m is the magnetic quantum number. $R_{n\ell}(r)$ can be written in terms of Laguerre polynomials and $Y_{\ell m}(\theta, \phi)$ are the spherical harmonic functions. In the case of isotropic effective masses, the eigenvalues $\mathcal{E}_r$ are given by:

$$\mathcal{E}_r = \mathcal{E}_g - \frac{\text{Ry}^*}{n^2}, \tag{6.133}$$

where the energy gap $\mathcal{E}_g$ represents the minimum energy of the continuum states and Ry* is the Rydberg constant for the exciton defined as:

$$\text{Ry}^* = \frac{m_r e^4}{2(4\pi\epsilon\hbar)^2} = 13.6\left(\frac{m_r}{m_0\epsilon_r^2}\right)\ \text{eV}. \tag{6.134}$$

Therefore the wavefunctions and energy values for an exciton are given by:

$$\psi_{n\ell m}(\mathbf{R}, \mathbf{r}) = \frac{1}{\sqrt{V}}\exp(i\mathbf{K}\cdot\mathbf{R})R_{n\ell}(r)Y_{\ell m}(\theta, \phi) \tag{6.135}$$

$$\mathcal{E}_{n\ell m} = \mathcal{E}_g + \frac{\hbar^2 K^2}{2M} - \frac{\text{Ry}^*}{n^2} \tag{6.136}$$

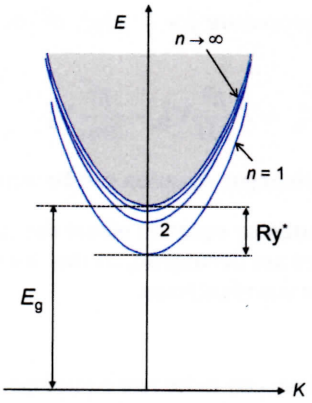

Figure 6.19: Energy states of a Wannier excitons. The bound states from $n = 1$ to $n = 3$ are clearly visible; the shaded area represents the continuum states.

The dispersion curves of a Wannier exciton are shown in Fig. 6.19. The exciton envelope functions $g_{n\ell m}(r)$ are the hydrogen-like functions. Therefore the ground-state exciton envelope function is given by:

$$g_{100}(r) = \frac{1}{\sqrt{\pi a_B^{*3}}} \exp\left(-\frac{r}{a_B^*}\right) \tag{6.137}$$

with

$$a_B^* = \left(\epsilon_r \frac{m_0}{m_r}\right) a_B \tag{6.138}$$

and a_B is the Bohr radius ($a_B = 0.53$ Å). Therefore, the interaction between the electron and hole in an exciton leads to the formation of discrete states below the conduction bandedge. The energy term $\hbar^2 K^2/2M$ corresponds to the kinetic energy of the free motion of the exciton CM. Above $\mathcal{E}_g + \frac{\hbar^2 K^2}{2M}$ there is a continuum of allowed energy states.

6.10.1 Absorption spectrum

We will now calculate how the presence of excitons modifies the absorption spectrum of a semiconductor. From a general point of view, the absorption spectrum can be calculated starting from the eigenfunctions and eigenenergies given in Eq. 6.135 and 6.136 and considering a suitable Hamiltonian describing the interaction between excitons and photons. The procedure is rather complex, therefore we will limit to mention the main final results.

It is possible to demonstrate that the imaginary part of the optical susceptibility can be written as:

$$\mathrm{Im}\chi(\omega) = \frac{e^2 |p_{vc}|^2 \sqrt{\mathrm{Ry}^*}}{2\epsilon_0 m^2 \omega^2} \left(\frac{2m_r}{\hbar^2}\right)^{3/2} \left[\sum_n \frac{4}{n^3} \delta\left(\Delta + \frac{1}{n^2}\right) + \frac{H(\Delta) e^{\pi/\sqrt{\Delta}}}{\sinh(\pi/\sqrt{\Delta})}\right] \tag{6.139}$$

which is called *3D Elliott formula*. In Eq. 6.139:

$$\Delta = \frac{\hbar\omega - \mathcal{E}_g}{\mathrm{Ry}^*} \tag{6.140}$$

and $H(\Delta)$ is the Heaviside step function. By using Eq. 5.51, $|p_{vc}|^2 = m^2\omega^2|\mu_{vc}|^2/e^2$, the 3D Elliott formula can be rewritten as:

$$\mathrm{Im}\chi(\omega) = \frac{|\mu_{vc}|^2\sqrt{\mathrm{Ry}^*}}{2\epsilon_0}\left(\frac{2m_r}{\hbar^2}\right)^{3/2}\left[\sum_n \frac{4}{n^3}\delta\left(\Delta + \frac{1}{n^2}\right) + \frac{H(\Delta)e^{\pi/\sqrt{\Delta}}}{\sinh(\pi/\sqrt{\Delta})}\right] \tag{6.141}$$

The absorption coefficient is given by:

$$\alpha = \frac{\omega}{nc}\mathrm{Im}\chi(\omega) \tag{6.142}$$

As we have already observed, the absorption spectrum is composed by a series of sharp lines with an amplitude rapidly decreasing upon increasing the quantum number n (the amplitude is proportional to n^{-3}) and by a continuum related to the ionized electron and hole particles.
It is instructive to consider the 3D Elliott formula when Coulomb effects become negligible, i.e., when excitonic effects are not present. In this situation the relative dielectric constant ϵ_r becomes very large so that Ry* becomes very small. We will assume $\mathrm{Ry}^* \to 0$, so that $\Delta \to \infty$:

$$\lim_{\Delta\to\infty} \frac{e^{\pi/\sqrt{\Delta}}}{\sinh(\pi/\sqrt{\Delta})} = \frac{\sqrt{\Delta}}{\pi} \tag{6.143}$$

Moreover we do not have to consider the contribution from bound exciton, so that the absorption spectrum given by Eqs. 6.142 and 6.141 is given by:

$$\begin{aligned}\alpha_{free}(\omega) &= \frac{\omega|\mu_{vc}|^2\sqrt{\mathrm{Ry}^*}}{2\epsilon_0 nc}\left(\frac{2m_r}{\hbar^2}\right)^{3/2}\frac{\sqrt{\Delta}}{\pi} = \\ &= \frac{\omega|\mu_{vc}|^2}{2\pi\epsilon_0 nc}\left(\frac{2m_r}{\hbar^2}\right)^{3/2}\sqrt{\hbar\omega - \mathcal{E}_g} = \\ &= \frac{\omega\pi}{nc\epsilon_0\hbar}|\mu_{vc}|^2\rho_j(\omega),\end{aligned} \tag{6.144}$$

which is the expression we have obtained in the case of bulk semiconductors assuming $f_v(\hbar\omega) - f_c(\hbar\omega) = 1$.
From Eqs. 6.141 and 6.142 the contribution to the absorption coefficient from the continuum states is given by:

$$\alpha_{cont} = \frac{\omega|\mu_{vc}|^2\sqrt{\mathrm{Ry}^*}}{2\epsilon_0 nc}\left(\frac{2m_r}{\hbar^2}\right)^{3/2}\frac{H(\Delta)e^{\pi/\sqrt{\Delta}}}{\sqrt{\Delta}\sinh(\pi/\sqrt{\Delta})} \tag{6.145}$$

which can be written in terms of $\alpha_{free}(\omega)$ as:

$$\alpha_{cont}(\omega) = \alpha_{free}(\omega)\frac{\pi}{\sqrt{\Delta}}\frac{e^{\pi/\sqrt{\Delta}}}{\sinh(\pi/\sqrt{\Delta})} = \alpha_{free}(\omega)S(\omega), \tag{6.146}$$

where $S(\omega)$ is called *Sommerfeld or Coulomb enhancement factor*. At the bandedge $\Delta \to 0$

$$\lim_{\Delta\to 0} \frac{e^{\pi/\sqrt{\Delta}}}{\sinh(\pi/\sqrt{\Delta})} = 2, \tag{6.147}$$

so that the Sommerfeld factor is approximately given by:

$$S(\omega) \to 2\frac{\pi}{\sqrt{\Delta}} \tag{6.148}$$

Therefore, at the bandedge the absorption does not go to zero following a square-root law as in the bulk case without excitonic effects, but it assumes a constant value given by:

$$\alpha(\Delta \to 0) = \frac{\omega|\mu_{vc}|^2\sqrt{\mathrm{Ry}^*}}{\epsilon_0 nc}\left(\frac{2m_r}{\hbar^2}\right)^{3/2} \tag{6.149}$$

Figure 6.20 shows a schematic bandedge absorption spectrum of a bulk semiconductor. It is evident that the presence of excitons does not only introduce discrete states

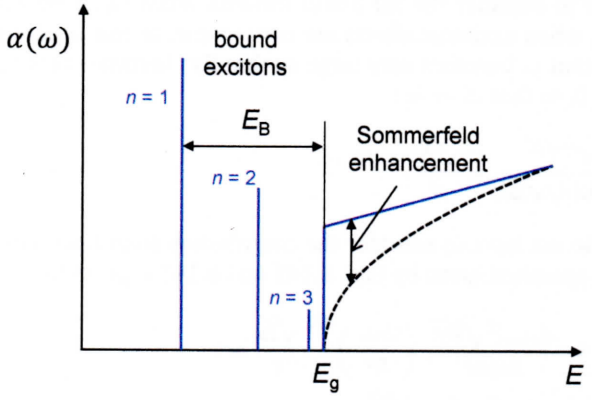

Figure 6.20: Very schematic plot of the absorption spectrum of a bulk semiconductor with (solid blue line) and without (dashed black line) excitonic effects.

in the gap, but significantly changes also the absorption in the continuum. Moreover, if the spectral broadening is taken into account (induced mainly by scattering of the electron-hole pairs with phonons) only a few bound states can be spectrally resolved, since the higher bound states merge together with the absorption of the ionized states. In real bulk semiconductors the absorption features related to the bound excitons can be experimentally observed at low temperatures in good-quality semiconductors.

6.10.2 Excitons in quantum wells

In the case of QWs the well thickness is typically smaller than the exciton Bohr radius in three-dimensional semiconductors; moreover the Coulomb potential in the direction perpendicular to the well does not play a significant role compared to the confinement potential. We have a kind of "two-dimensional" hydrogen atom, which introduces discrete energy levels in the gap, whose dispersion relations are given by the following expression:

$$\mathcal{E}_{nK} = \mathcal{E}_g + \mathcal{E}_{1c} + \mathcal{E}_{1v} + \frac{\hbar^2 K^2}{2M} - \frac{\mathrm{Ry}^*}{\left(n - \frac{1}{2}\right)^2} \tag{6.150}$$

The fundamental binding energy of excitons in QWs is four times the corresponding binding energy in the three-dimensional case, indeed $\mathcal{E}_B = 4\text{Ry}^*$.

It is possible to demonstrate that the imaginary part of the optical susceptibility describing the excitonic absorption in a QW with thickness L is given by the *2D Elliott formula*

$$\text{Im}\chi(\omega) = \frac{e^2|p_{vc}|^2 m_r}{\epsilon_0 m^2 \hbar^2 \omega^2 L}\left\{\sum_n \frac{4}{\left(n+\frac{1}{2}\right)^3}\delta\left[\Delta + \frac{1}{\left(n+\frac{1}{2}\right)^2}\right] + \frac{H(\Delta)e^{\pi/\sqrt{\Delta}}}{\cosh(\pi/\sqrt{\Delta})}\right\} \quad (6.151)$$

where:

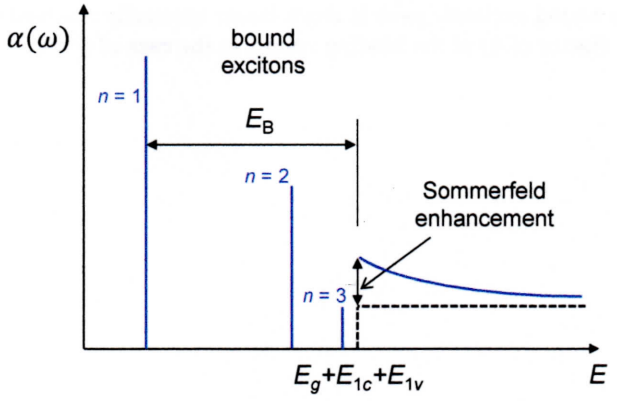

Figure 6.21: Very schematic plot of the absorption spectrum of a QW with (solid blue line) and without (dashed black line) excitonic effects.

$$\Delta = \frac{\hbar\omega - (\mathcal{E}_g + \mathcal{E}_{1c} + \mathcal{E}_{1v})}{\text{Ry}^*} \quad (6.152)$$

Equation 6.151 can be also written as:

$$\text{Im}\chi(\omega) = \frac{|\mu_{vc}|^2 m_r}{\epsilon_0 \hbar^2 L}\left\{\sum_n \frac{4}{\left(n+\frac{1}{2}\right)^3}\delta\left[\Delta + \frac{1}{\left(n+\frac{1}{2}\right)^2}\right] + \frac{H(\Delta)e^{\pi/\sqrt{\Delta}}}{\cosh(\pi/\sqrt{\Delta})}\right\} \quad (6.153)$$

and the absorption spectrum can be calculated by using Eq. 6.142. Also in the case of QWs, in the limit of completely negligible Coulomb interaction (limit of independent electrons, $\Delta \to \infty$) the absorption coefficient given by Eq. 6.153 turns out to be given by what we have already calculated without excitonic effects. Indeed in this case

$$\lim_{\Delta\to\infty} \frac{e^{\pi/\sqrt{\Delta}}}{\cosh(\pi/\sqrt{\Delta})} = 1 \quad (6.154)$$

so that the absorption spectrum is given by:

$$\alpha_{free}(\omega) = \frac{\omega|\mu_{vc}|^2 m_r}{cn\epsilon_0\hbar^2 L} H[\hbar\omega - (\mathcal{E}_g + \mathcal{E}_{1c} + \mathcal{E}_{1v})] \quad (6.155)$$

which is the expression we have obtained in the case of free charges assuming $f_v(\hbar\omega) - f_c(\hbar\omega) = 1$.
Also in this case we can write the continuum contribution as:

$$\alpha_{cont}(\omega) = \alpha_{free}(\omega)\frac{e^{\pi/\sqrt{\Delta}}}{\cosh(\pi/\sqrt{\Delta})} = \alpha_{free}(\omega)S_{free}(\omega) \tag{6.156}$$

Since at the bandedge ($\Delta \to 0$) the Sommerfeld factor is 2, the absorption at the band edge is twice the free-carrier absorption. Figure 6.21 shows a schematic bandedge absorption spectrum in a QW. As in the bulk case, considering the spectral broadening of the sharp excitonic states in the gap, the absorption in the higher band states merges with the absorption of the ionized states. In contrast to the bulk case, the fundamental $1s$ bound excitonic peak is much better spectrally resolved due to the strong increase (factor of 4) of the binding energy in the case of QWs.

7. Light Emitting Diodes

7.1 Basic concepts

In this chapter the generation of light by the process of electroluminescence in a semiconductor will be considered. This phenomenon was first reported in 1907 by Henry Joseph Round, who observed the emission of visible light from carborundum (i.e., an impure form of polycristalline SiC) when a potential was applied between two points of the crystal. In 1951 Lehovec and co-workers explained the electroluminescence in SiC as the result of injection of carriers across a junction followed by radiative recombination of the electron-hole pairs. The first demonstration of a direct bandgap visible (GaAsP) light-emitting diode (LED) was reported by Holonyak and Bevaqua in 1962. Towards the end of the 1960s the mass production of low-cost GaAsP red LEDs began (first in Monsanto and then at Hewlett-Packard).

A LED is essentially a $p-n$ junction diode under forward bias, as shown in Fig. 7.1 in the case of a homo-junction. We will first describe this simple case although homo-junctions are not used in real devices (as we will discuss in this chapter). Under forward bias electrons from the $n-$side of the junction and holes from the $p-$side cross the depletion layer thus becoming minority carriers. These excess minority carriers diffuse in the semiconductor (minority electrons diffuse in the $p-$type semiconductor, while minority holes diffuse in the $n-$type semiconductor) and recombine with majority carriers. For example, the excess minority electron concentration in the $p-$type portion varies along the semiconductor as:

$$\Delta n(x) = \Delta n(0) \exp(-x/L_n), \tag{7.1}$$

where L_n is the diffusion length of electrons, defined as the average distance a minority carrier diffuses before recombination. The diffusion length is given by:

$$L_n = \sqrt{D_n \tau_n}, \qquad L_p = \sqrt{D_p \tau_p} \tag{7.2}$$

where D_n and D_p are the diffusion constants for electrons and holes, respectively, τ_n and τ_p are the electron and hole carrier lifetimes. In general the diffusion length in a

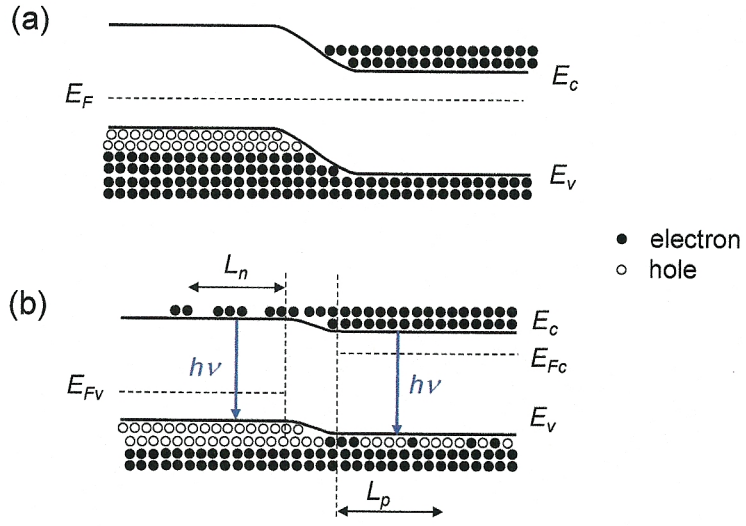

Figure 7.1: $p-n$ homojunction under zero bias (a) and under forward bias (b).

semiconductor is of the order of a few micrometers. As we have discussed in Chapter 5, electron-hole recombination can occur by means of radiative and non-radiative processes: a LED is designed to maximize the fraction of radiative events. Of course, if the $p-n$ junction is under reverse bias, no carriers are injected across the junction and no photons can be generated. The current-to-voltage characteristic for a diode can be written as follows:

$$I = I_s \left[\exp\left(\frac{eV}{k_B T}\right) - 1\right] \simeq I_s \exp\left(\frac{eV}{k_B T}\right), \tag{7.3}$$

where I_s is the reverse saturation current and V is the voltage applied to the diode.

In the following we will recall a few important concepts about radiative and non-radiative recombination processes we have already discussed in Chapter 5. Equation 5.67 gives the spontaneous emission rate, $R_{sp}(\hbar\omega)$ per unit time, volume and energy (for photons with energy $\hbar\omega$):

$$R_{sp}(\hbar\omega) = \frac{1}{\tau_r}\rho_j(\hbar\omega) f_c(\hbar\omega)[1 - f_v(\hbar\omega)], \tag{7.4}$$

where τ_r is the spontaneous emission lifetime. In general, in the case of LEDs the excess carrier concentration is rather small so that the quasi-Fermi levels are in the gap and quite far from the bandedges, as shown in Fig. 7.1(b). Therefore the Fermi-Dirac distribution function can be approximated by the Boltzmann distribution function. With this approximation the spontaneous emission rate, which gives the spontaneous emission lineshape, can be written as (see Eq. 5.77):

$$R_{sp}(\hbar\omega) = \Re_{sp}\,(\hbar\omega - \mathcal{E}_g)^{1/2} \exp\left(-\frac{\hbar\omega - \mathcal{E}_g}{k_B T}\right), \tag{7.5}$$

where $\Re_{sp}$ is given by the following expression (see Eq. 5.76):

$$\Re_{sp} = \frac{1}{2\pi^2\tau_r}\left(\frac{2m_r}{\hbar^2}\right)^{3/2} \exp\left(\frac{\Delta\mathcal{E}_F - \mathcal{E}_g}{k_B T}\right), \tag{7.6}$$

where $\Delta\mathcal{E}_F$ is the difference between the quasi-Fermi energies: $\Delta\mathcal{E}_F = \mathcal{E}_{Fc} - \mathcal{E}_{Fv}$. In Chapter 5 it has been demonstrated that the photon energy corresponding to the peak value of the spontaneous emission lineshape is $\hbar\omega = \mathcal{E}_g + k_B T/2$ with a full-width at half-maximum $\Delta\mathcal{E} = 1.8 k_B T$, which can be written in terms of the emission wavelength as follows (see Ex. 5.4):

$$\Delta\lambda = 1.8 k_B T \frac{\lambda^2}{hc}. \tag{7.7}$$

For example, an AlGaAs/GaAs LED emitting at 625 nm has a (theoretical) linewidth $\Delta\mathcal{E} = 46$ meV (at $T = 300$ K), corresponding to $\Delta\lambda \simeq 15$ nm. Figure 7.2 shows the spectral distribution of various LEDs with emission from the violet to the near-infrared.

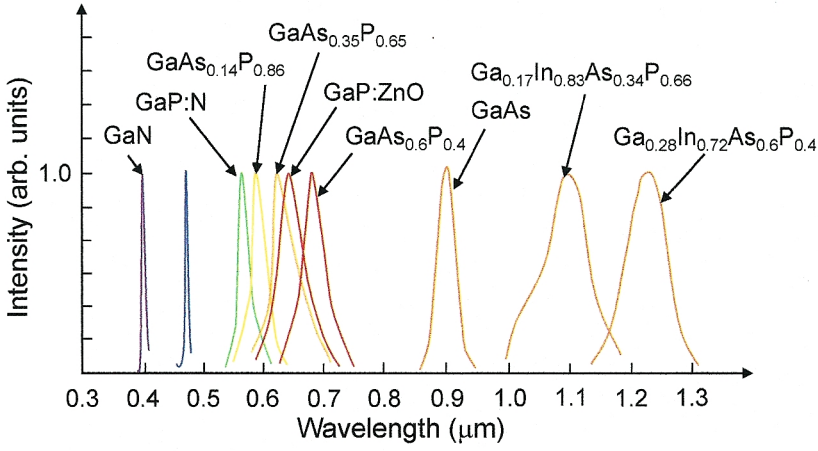

Figure 7.2: Emission spectra of various LEDs from violet to near-infrared.

We have then calculated the total radiative recombination rate, $R_{sp,T}$, for photons of all energies (see Eq. 5.91) in the case of non-degenerate semiconductors, so that Maxwell-Boltzmann statistics can be applied. This is the typical situation in the case of LEDs, so that the radiative recombination rate can be written as:

$$R_{sp,T} = Bnp = \frac{\Delta n}{\tau_{rad}}, \tag{7.8}$$

where Δn is the non-equilibrium carrier density in the junction region of the LED, B is the bimolecular recombination coefficient, given by Eq. 5.92 and τ_{rad} is the radiative recombination time. In the following we will assume $\Delta n = \Delta p = n$. We can calculate the total photon flux, Φ, per second as:

$$\Phi = R_{sp,T} \frac{V_{jun}}{S} = \frac{n}{\tau_{rad}} d \tag{7.9}$$

where V_{jun} is the volume of the junction ($V_{jun} = Sd$, with d the effective thickness of the junction, where electron-hole recombination takes place).

The non-radiative recombination processes include the Shockley-Read-Hall (SRH) process and the Auger recombination, which depend on the minority carrier concentration, n (electrons in a $p-$type layer, holes in a $n-$type layer). We have seen

that the SRH process increases linearly, $A_{nr}n$ (see Eq. 5.135), whereas the Auger recombination process in the high-excitation limit increases as the cube of n, Cn^3 (or, more generally, $f(n)$), since it is a three-particle process. A_{nr} and C are constants. Therefore, the non-radiative lifetime, τ_{nr}, can be written as:

$$\frac{1}{\tau_{nr}} = A_{nr} + Cn^2. \tag{7.10}$$

The total recombination time, τ (see Eq. 5.189) is given by:

$$\frac{1}{\tau} = \frac{1}{\tau_{rad}} + \frac{1}{\tau_{nr}} = A_{nr} + Bn + Cn^2, \tag{7.11}$$

The probability of radiative recombination, called *internal quantum efficiency*, η_i (Eq. 5.190), is given by:

$$\eta_{int} = \frac{1/\tau_{rad}}{1/\tau_{rad} + 1/\tau_{nr}} = \frac{\tau}{\tau_{rad}}, \tag{7.12}$$

i.e., it is defined as the ratio between the probability of radiative recombination and the total recombination probability. If J_e and J_h are the current densities for electrons and holes, the electron and hole fluxes into the junction region are given by J_e/e and J_h/e. To preserve charge neutrality we must have $J_e = J_h = J$. In stationary state conditions the input flux of carriers which are flowing through the surface S of the junction equals the total recombination rate, R_t, in the volume V_{jun}, so that:

$$\frac{J}{ed} = A_{nr}n + Bn^2 + Cn^3 = \frac{n}{\tau}. \tag{7.13}$$

By using this last expression in Eq. 7.9, the total photon flux can be written as:

$$\Phi = \eta_{int}\frac{J}{e}. \tag{7.14}$$

From this relation the physical meaning of η_i is clear: it represents the ratio between the number of photons generated inside the LED and the total number of injected electrons.

7.2 Double heterostructure LEDs

At low carrier densities the non-radiative SRH term (A_{nr}) becomes the dominant term, thus leading to a low internal quantum efficiency. This is the situation in the case of a $p-n$ homojunction LED, shown in Fig. 7.1. In this case the electrons diffuse into the $p-$type layer and the holes diffuse into the $n-$type layer. The effective thickness d of the junction is given by $L_n + L_p$ and the diffusion length for minority carriers is typically in a range between 1 μm and 10 μm, therefore the minority carriers diffuse over a large region thus reducing their density. In order to significantly decrease the effective thickness d thus increasing the minority carrier density, LEDs are typically based on double heterostructures, in which the thickness of the active layer (i.e., the layer where radiative recombinations occur) is typically around 0.2-0.3 μm. Therefore, at the same value of the current density the carrier concentration increases, thus leading to a significant increase of the probability of radiative recombination (Bn^2) and of the LED efficiency.

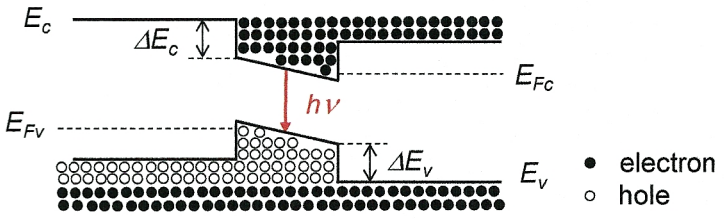

Figure 7.3: Schematic energy bands of a double-heterostructure LED under forward bias.

A double heterostructure is formed by an active region, where radiative recombination occurs, and two confinement layers cladding the active region, characterized by a larger energy gap, as shown schematically in Fig. 7.3. In order to obtain an effective carrier confinement the band discontinuities $\Delta\mathcal{E}_c$ and $\Delta\mathcal{E}_v$ in the conduction and valence band, respectively, must be larger than the thermal energy k_BT, to block carrier leakage into the cladding layers. Double heterostructures can be employed both in the case of bulk semiconductors and in the case of quantum wells, where additional carrier localization is achieved in the well region.
An additional important advantage of double heterostructures with respect to homojunctions is that the photons emitted in the active region are not absorbed by the confinement semiconductors, which have a larger energy gap. Since the confinement regions are rather thin, they are almost totally transparent to the emitted radiation. Moreover, under typical carrier injection levels, also the active region is almost transparent to the emitted radiation. Indeed, as shown in Fig. 7.4, due to carrier injection in the active layer, the quasi-Fermi levels in the conduction and valence bands are within the bands, so that all states close to the conduction band edge are populated by electrons and transitions from the valence band to these states cannot occur, thus leading to a shift of the absorption edge to higher energies: this effect is called *Burstein-Moss shift*. It is therefore evident why homojunctions are not used in LEDs.

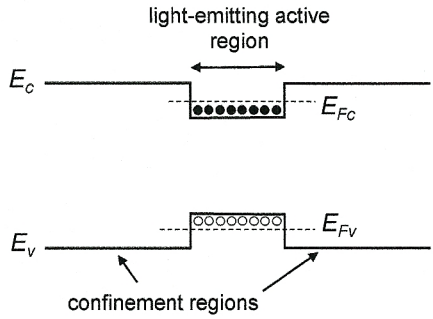

Figure 7.4: Qualitative interpretation of the Burnstein-Moss shift.

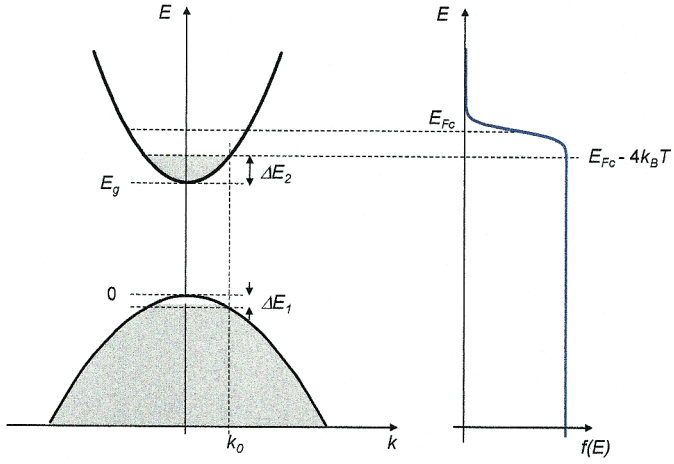

Figure 7.5: Dispersion curves in the conduction and valence band of a bulk semiconductor; the shaded areas represent occupied states. On the right the Fermi-Dirac distribution function is reported. It is assumed that all levels with energy below $\mathcal{E}_{Fc} - 4k_BT$ are populated.

7.2.1 Burstein-Moss shift.

If we assume that all levels in the conduction band with energy below $\mathcal{E}_F - 4k_BT$ are populated, the minimum energy of an absorbed photon is given by (see Fig. 7.5):

$$(\hbar\omega)_{min} = \mathcal{E}_g + \Delta\mathcal{E}_1 + \Delta\mathcal{E}_2. \tag{7.15}$$

The corresponding k–value is k_0 as displayed in Fig. 7.5. At $k = k_0$ the energy $\mathcal{E}_2$ of the corresponding state in the conduction band is:

$$\mathcal{E}_2 = \mathcal{E}_g + \frac{\hbar^2 k_0^2}{2m_c} = \mathcal{E}_{Fc} - 4k_BT, \tag{7.16}$$

so that

$$\frac{\hbar^2 k_0^2}{2m_c} = \mathcal{E}_{Fc} - 4k_BT - \mathcal{E}_g = \Delta\mathcal{E}_2. \tag{7.17}$$

The energy $\mathcal{E}_1$ of the corresponding state in the valence band is given by:

$$\mathcal{E}_1 = -\frac{\hbar^2 k_0^2}{2m_v} = -\frac{m_c}{m_v}(\mathcal{E}_{Fc} - 4k_BT - \mathcal{E}_g) = -\Delta\mathcal{E}_1. \tag{7.18}$$

The Burnstein-Moss shift of the absorption edge is given by:

$$\Delta\mathcal{E} = (\hbar\omega)_{min} - \mathcal{E}_g = \Delta\mathcal{E}_1 + \Delta\mathcal{E}_2. \tag{7.19}$$

From Eqs. 7.17-7.19 we have:

$$\Delta\mathcal{E} = (\mathcal{E}_{Fc} - 4k_BT - \mathcal{E}_g)\left(1 + \frac{m_c}{m_v}\right). \tag{7.20}$$

As we have seen in Sect. 2.3, the relationship between the electron density and Fermi energy is given by:

$$n = N_c F_{1/2}(y), \tag{7.21}$$

where:

$$y = \frac{\mathcal{E}_{Fc} - \mathcal{E}_c}{k_BT} = \frac{\mathcal{E}_{Fc} - \mathcal{E}_g}{k_BT}. \tag{7.22}$$

In the case we are considering y is positive and can be assumed $\gg 1$, so that the Fermi function can be approximated as in Eq. 2.44, so that:

$$n = \frac{4N_c}{3\sqrt{\pi}}\left(\frac{\mathcal{E}_{Fc} - \mathcal{E}_g}{k_BT}\right)^{3/2}. \tag{7.23}$$

From this expression it is possible to calculate $\mathcal{E}_{Fc} - \mathcal{E}_g$:

$$\mathcal{E}_{Fc} - \mathcal{E}_g = k_BT\left(\frac{3\sqrt{\pi}}{4}\frac{n}{N_c}\right)^{2/3}. \tag{7.24}$$

A very important device parameter which strongly influences the LED efficiency is the thickness of the active region. Typical thickness is of the order of a few tenth of micrometer in the case of bulk active regions, as shown in Fig. 7.6 in the case of an AlGaInP double heterostructure LED. The decrease of LED efficiency upon increasing

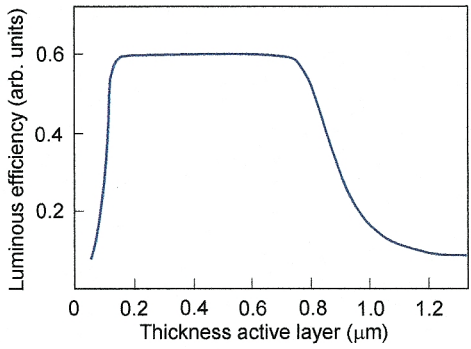

Figure 7.6: Luminous efficiency of an AlGaInP double heterostructure LED vs the thickness of the active layer with emission at 565 nm. Adapted from Sugawara *et al.*, US Papent 5, 153, 889 (1992).

the thickness of the active region above ~ 1 μm can be easily understood: carrier confinement induced by the double heterostructure is less effective and when the thickness becomes comparable or larger than the carrier diffusion length, double

heterostructures do not present any advantage compared to homojunctions. If the thickness of the active region is decreased too much at moderately high levels of carrier injection there is a significant carrier overflow, thus leading to an increase of the losses. An estimation of the carrier concentration required to have carrier overflow can be obtained as follows.

7.2.2 Carrier overflow

We will consider the conduction band, but the same considerations can be repeated also for the valence band. Under steady-state conditions the injection of carriers equals the rate of carrier recombination. Therefore, assuming only radiative recombination:

$$\frac{J}{ed} = Bnp. \tag{7.25}$$

For high injection density $n = p$ and from 7.25:

$$n = \sqrt{\frac{J}{eBd}}. \tag{7.26}$$

The condition for carrier overflow is reached when the separation between the quasi-Fermi level and the bottom of the conduction band, $\mathcal{E}_{Fc} - \mathcal{E}_c$, equals the height of the potential barrier between the active region and the cladding layers:

$$\mathcal{E}_{Fc} - \mathcal{E}_c = \mathcal{E}_B. \tag{7.27}$$

The net barrier height includes residual band bending under forward bias and the conduction band discontinuity $\Delta\mathcal{E}_c$. By using Eq. 7.24 we can write:

$$k_B T \left(\frac{3\sqrt{\pi}}{4}\frac{n}{N_c}\right)^{2/3} = \mathcal{E}_B \tag{7.28}$$

so that:

$$n = \frac{4}{3\sqrt{\pi}} N_c \left(\frac{\mathcal{E}_B}{k_B T}\right)^{3/2}. \tag{7.29}$$

The corresponding overflow current density can be written as:

$$J = eBd\left(\frac{4}{3\sqrt{\pi}}N_c\right)^2 \left(\frac{\mathcal{E}_B}{k_B T}\right)^3. \tag{7.30}$$

■

Exercise 7.1 Calculate the current density required for carrier overflow at room temperature in a GaAs double heterostructure assuming a barrier height $\Delta\mathcal{E}_c = 180$ meV and a thickness $d = 0.1$ μm. The bimolecular recombination coefficient is $B = 1.7 \times 10^{-10}$ cm^{3}s^{-1}.

In Ex. 2.3 the effective density of states in the conduction band of GaAs has

been calculated: $N_c = 4.4 \times 10^{17}$ cm^{-3}. From Eq. 7.31 one obtains:

$$J = eBd\left(\frac{4}{3\sqrt{\pi}}N_c\right)^2\left(\frac{\mathcal{E}_B}{k_BT}\right)^3 = 9.9 \times 10^3 \text{ A/cm}^2. \tag{7.31}$$

The current density for carrier overflow decreases with the thickness, d, of the active region since less carriers are required to completely fill the active region. When carrier overflow is activated, an increase of carrier injection does not lead to an increase of the emitted radiation, thus leading to a saturation of the emitted light intensity upon increasing the diode current (i.e., it is an effective loss process). This effect is important also in the case of heterostructures with quantum wells.

7.3 Carrier leakage over barrier

At lower levels of carrier injection, far from carrier overflow, another loss mechanism is still active, related to the energy distribution of the carriers in the conduction and valence band. Typical heights of the potential barriers confining carriers in the active region ($\Delta\mathcal{E}_c$ and $\Delta\mathcal{E}_v$) are of the order of a few hundreds of meV, much larger than the thermal energy k_BT. Due to the shape of the Fermi-Dirac distribution function, a few carriers have energy larger than the height of the confining barriers: these carriers can escape from the active region and diffuse in the barrier cladding region. This situation is schematically displayed in Fig. 7.7, in the case of electrons in the conduction band. The density of electrons in the conduction band with energy larger

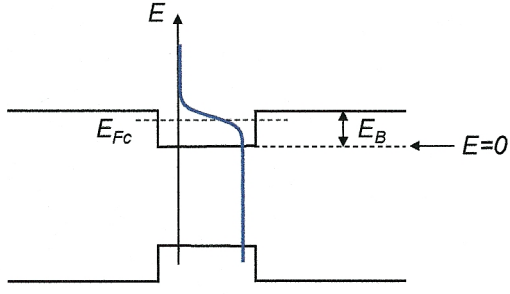

Figure 7.7: Carrier leakage over barrier: due to the shape of the Fermi-Dirac distribution function (blue curve) the high-energy electrons can escape from the active region and diffuse in the barrier cladding region.

than the effective barrier height $\mathcal{E}_B$ is given by:

$$n_B = \int_{\mathcal{E}_B}^{\infty} \rho_c(\mathcal{E}) f_c(\mathcal{E}) d\mathcal{E}, \tag{7.32}$$

Since we are at moderate charge injection and we are considering electrons with energy much larger than the Fermi energy, the Fermi-Dirac distribution function can be well approximated by the Boltzmann distribution:

$$f(\mathcal{E}) \simeq \exp\left(-\frac{\mathcal{E} - \mathcal{E}_{Fc}}{k_BT}\right), \tag{7.33}$$

so that the electron density (Eq. 7.32) is given by (see Eq. 2.47):

$$n_B = N_c \exp\left(-\frac{\mathcal{E}_B - \mathcal{E}_{Fc}}{k_B T}\right) \tag{7.34}$$

A fraction of the electron population with energy above the barrier is reflected at the interface between the active region and the cladding region. Another fraction diffuse into the cladding layer. Note that, since the velocities of the high-energy electrons have random orientations, only a portion of these electrons are moving towards the cladding layer. In addition, a fraction of the electrons which enter the cladding material can diffuse back towards the active region. Therefore it is not trivial to calculate the effective carrier leakage over the barrier.

To calculate the electron leakage current we can assume that the electrons diffuse into the $p-$type cladding layer as minority carriers, so that the variation of the electron density in this region can be written as:

$$n_c(x) = n_c(0) \exp\left(-\frac{x}{L_n}\right) = n_B \exp\left(-\frac{x}{L_n}\right). \tag{7.35}$$

where $L_n = \sqrt{D_n \tau_n}$ is the diffusion length, D_n is the diffusion constant and τ_n is the minority carrier lifetime in the $p-$cladding medium. We have assumed an $x-$axis with origin at the active region-cladding interface and $n_c(0) = n_B$, given by Eq. 7.34. The corresponding diffusion current is given by:

$$J_n|_{x=0} = -eD_n \frac{dn}{dx} = eD_n \frac{n_B}{L_n} = e\frac{L_n}{\tau_n} n_B. \tag{7.36}$$

Exercise 7.2 Consider a GaAs/$Al_{0.3}Ga_{0.7}As$ double heterostructure with an injected electron density in the GaAs active region $n = 2.86 \times 10^{18}$ cm^{-3} ($T = 300$ K). The barrier height between the active and the cladding regions is 200 meV. The electron mobility in the cladding region is $\mu_n = 2300$ cm^2/Vs and the minority carrier lifetime is 14 ns. Calculate the leakage current density.

In Ex. 2.3 the position of the quasi-Fermi level in the conduction band of GaAs with an electron density $n = 2.86 \times 10^{18}$ has been calculated:

$$\mathcal{E}_{Fc} - \mathcal{E}_c = 109 \text{ meV}.$$

The density of electrons with energy above the barrier can be calculated by using Eq. 7.34:

$$n_B = N_c \exp\left(-\frac{\mathcal{E}_B - \mathcal{E}_{Fc}}{k_B T}\right) = 1.3 \times 10^{16} \text{ cm}^{-3}.$$

The diffusion constant is given by

$$D_n = \mu_n \frac{k_B T}{e} = 59.5 \text{ cm}^2/\text{s},$$

and the diffusion length is:

$$L_n = \sqrt{D_n \tau_n} = 9.1 \ \mu\text{m}.$$

The leakage current density is given by Eq. 7.36:

$$L_n = e\frac{D_n}{L_n} n_B = 136 \text{ A/cm}^2.$$

This result shows that the leakage current represents a significant loss mechanism.

■

Note that, as a consequence of the exponential dependence of n_B on temperature, the leakage current increases rapidly with temperature. Moreover, a reduced minority carrier lifetime caused by high defect density determines an increase of the leakage current.

To reduce the carrier leakage from the active region, which can be an important loss mechanism, in particular in the case of double heterostructures with low barrier heights at the interface between the active region and the confinement materials, carrier-blocking layers can be used. In particular electron blocking layers are employed, due to the typically larger diffusion constant of electrons in comparison with holes in III-V semiconductors. A blocking layer is formed by a thin semiconductor layer with high gap, placed at the interface between active and confining regions, as schematically shown in Fig. 7.8.

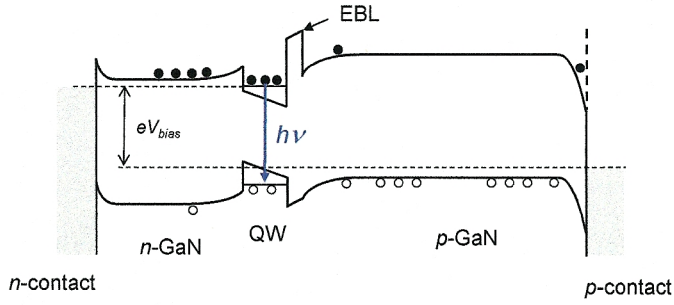

Figure 7.8: Energy-level diagram of a GaN quantum well (QW) LED showing an electron blocking layer (EBL). V_{bias} is the forward bias voltage applied to the device.

Figure 7.8 also shows the band bending at the interfaces between different semiconductors, whose origin can be easily explained. Let us consider the interface between two semiconductors with different energy gaps. Carriers in the semiconductor with larger energy gap diffuse towards the semiconductor with smaller gap, filling states with lower energy. Let us consider the electron transfer in the conduction band: a charge separation is created at the interface with a depletion region with positive charge in the large-gap semiconductor and a negatively charged electron accumulation layer in the small-gap semiconductor. This charge separation leads to the generation of an electric field, which causes a bending of the band, as shown in Fig. 7.8. The potential energy discontinuities at the interfaces, which can cause an increase of the resistance to current flow, can be greatly reduced by using graded heterostructures, i.e., heterostructures where the semiconductor composition does not change abruptly. For example, in a $p-n$ junction n–$Al_{0.3}Ga_{0.7}As$/p–GaAs graded heterostructure, $Al_xGa_{1-x}As$ is graded from $x = 0$ to $x = 0.3$ over a short distance.

7.4 External efficiency of a LED

So far we have discussed the mechanisms leading to an increase or a decrease of the internal quantum efficiency of a LED. In a real device not all the photons generated in the active region are emitted. There are a number of different processes leading to a reduction of the number of photons at the output of the LED. Photons can be reabsorbed during their travel inside the semiconductor materials (this loss process can be minimized by using properly designed heterostructures). Photons incident on metallic electrodes are absorbed by the metal. Another (large) fraction of photons cannot exit from the LED due to total internal reflection. For this reason it is important to define a transmission efficiency, η_t, as the ratio between the number of photons emitted into the free space per second and the number of photons emitted from the active region per second. The product of the transmission efficiency and the internal quantum efficiency gives the ratio between the number of photons emitted into the free space per second and the number of electrons injected into the LED per second, which is called *external quantum efficiency*, η_{ext}:

$$\eta_{ext} = \eta_t \eta_{int}. \tag{7.37}$$

With this definition, the external photon flux is given by (see Eq. 7.14 for the photon flux inside the LED):

$$\Phi_{ext} = \eta_{ext} \frac{J}{e}, \tag{7.38}$$

The output power, P_{ext}, can be calculated by multiplying the external photon flux by the surface S and the energy of the emitted photons:

$$P_{ext} = \Phi_{ext} S h\nu = \eta_{ext} \frac{I}{e} h\nu. \tag{7.39}$$

The *responsivity* is defined as the ratio between the emitted optical power and the injected current:

$$R = \frac{P_{ext}}{I} = \eta_{ext} \frac{h\nu}{e}. \tag{7.40}$$

The responsivity in W/A can be written as follows, when the emission wavelength is expressed in μm:

$$R = \eta_{ext} \frac{1.24}{\lambda_0}. \tag{7.41}$$

The linear dependence of the output power vs input current is valid for relatively small values of current (typically a few tens of mA). Upon increasing the current, efficiency starts to decrease, as previously mentioned (current droop).

The power efficiency, or *wall-plug efficiency*, is defined as the ratio between the emitted optical power and the electrical power provided to the LED:

$$\eta_{wp} = \frac{P_{ext}}{IV}, \tag{7.42}$$

here V is the voltage drop across the device.

As already mentioned, a major mechanism which strongly reduces the number of photons emitted into the free space is due to total internal reflection at the

semiconductor-air interface. It is possible to calculate an approximated value of the fraction of the radiation which can exit from the interface between two materials with refractive index n_1 and n_2 ($n_1 > n_2$). We will assume an isotropic radiation source in the medium with refractive index n_1 and, as a first-order approximation, we will consider a constant transmission at the interface between the two media: $T(\theta) = T(0)$, where $T(0)$ is the normal incidence transmission. The light rays outside a cone defined by the critical angle θ_c corresponding to the condition of total internal reflection are completely reflected. θ_c can be easily calculated by using the Snell's law:

$$n_1 \sin\theta_c = n_2, \qquad \theta_c = \arcsin\frac{n_2}{n_1}. \tag{7.43}$$

Since the refractive index of a semiconductor is typically quite high ($n_1 \gg n_2$) we can use the following approximation:

$$\sin\theta_c \simeq \theta_c, \qquad \theta_c = \frac{n_2}{n_1}. \tag{7.44}$$

Indeed, if $n_1 = 3.4$ and $n_2 = 1$, $\theta_c = 17°$. The light rays which can escape from the

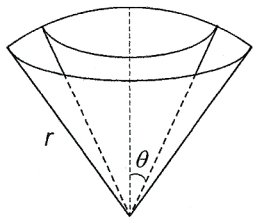

Figure 7.9: Calculation of the solid angle corresponding to the light escape cone.

semiconductor are those inside the light escape cone with semi-aperture θ_c, as shown in Fig. 7.9. The corresponding solid angle Ω_c is given by:

$$\Omega_c = \int_0^{\theta_c} \frac{2\pi r \sin\theta \; r d\theta}{r^2} = 2\pi \int_0^{\theta_c} \sin\theta \; d\theta = 2\pi(1 - \cos\theta_c). \tag{7.45}$$

For small angles θ_c, $\cos\theta_c \simeq 1 - \theta_c^2/2$, so that:

$$\Omega_c \simeq \pi\theta_c^2 = \pi\left(\frac{n_2}{n_1}\right)^2. \tag{7.46}$$

The power fraction emitted inside the light escape cone assuming an isotropic source is given by $\Omega_c/4\pi$, so that the power that can escape from the semiconductor is given by:

$$P_{ext} = P_{source} T(0) \frac{\Omega_c}{4\pi}, \tag{7.47}$$

where:

$$T(0) = 1 - \left(\frac{n_1 - n_2}{n_1 + n_2}\right)^2 \tag{7.48}$$

and the extraction efficiency related to total internal reflection is:

$$F = \frac{P_{ext}}{P_{source}} \approx \frac{1}{4}\left[1 - \left(\frac{n_1 - n_2}{n_1 + n_2}\right)^2\right]\left(\frac{n_2}{n_1}\right)^2 \tag{7.49}$$

A simple way to increase the critical angle and therefore also the fraction of emitted power is to encapsulate the LED in a plastic material with a refractive index larger than 1.

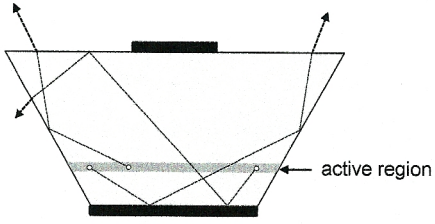

Figure 7.10: LED with truncated inverted pyramid (TIP) structure.

The light rays emitted outside the light escape cone are trapped inside the semiconductor and most of them are absorbed by the substrate. In order to avoid this problem a LED should have hemispherical shape with a point-like source in the center so that all the emitted light rays are normal to the semiconductor-air interface and can escape from the device (only Fresnel reflection must be considered at the interface, if the surface of the sphere does not have an anti-reflection coating). The problem of this structure is the fabrication cost, related to the non-planar geometry. A shaping technology more suitable for mass production, which has proven to increase LED efficiency, consists in the fabrication of LEDs in the shape of truncated inverted pyramid (TIP), which can redirect towards the exit axis side-emitted rays, as schematically shown in Fig. 7.10. An increase in efficiency of a factor of two is possible by using this particular geometry.

7.5 Emission pattern of a LED

The emission at the output of a LED is in general non-isotropic even though light inside the semiconductor can be assumed as isotropic. Let us consider a planar LED with high refractive index n_1 placed in air (refractive index n_2). As in the previous section, we will further assume an isotropic point source placed in the semiconductor, quite close to the semiconductor-air interface, as shown in Fig. 7.11. Moreover, a constant transmission is assumed, independent of the incidence angle. The photon flux emitted from the point source with directions between θ_1 and $\theta_1 + d\theta_1$ (i.e., within the solid angle $d\Theta_1$ delimited by two conical surfaces with semi-aperture θ_1 and $\theta_1 + d\theta_1$) is given by:

$$d\Phi_1(\theta_1) = I_1(\theta_1) 2\pi \sin\theta_1 \, d\theta_1 = I_1(0) 2\pi \sin\theta_1 \, d\theta_1 \tag{7.50}$$

where I is the luminous intensity (i.e., the luminous power per unit solid angle), which does not depend on θ_1 since the source is isotropic. The photon flux emerging in the second medium in the solid angle corresponding to $d\theta_2$ is:

$$d\Phi_2(\theta_2) = T(\theta_1) d\Phi_1(\theta_1) \tag{7.51}$$

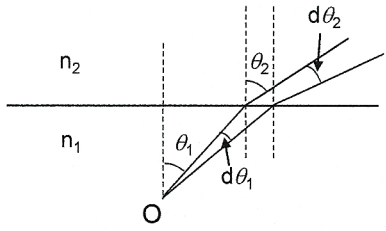

Figure 7.11: An isotropic point source placed in O in the medium with refractive index n_1 forms a Lambertian source in the medium with refractive index n_2.

We will assume $T(\theta_1) = T(\theta = 0) = T_0$. The photon flux $d\Phi_2$ can be written in terms of the luminous intensity in the region outside the LED:

$$d\Phi_2(\theta_2) = I_2(\theta_2) 2\pi \sin\theta_2 \, d\theta_2 \tag{7.52}$$

so that:

$$I_2(\theta_2) = T_0 I_1(0) \frac{\sin\theta_1 d\theta_1}{\sin\theta_2 d\theta_2} \tag{7.53}$$

From Snell's law:

$$n_1 \sin\theta_1 = n_2 \sin\theta_2, \qquad n_1 \cos\theta_1 d\theta_1 = n_2 \cos\theta_2 d\theta_2 \tag{7.54}$$

so that Eq. 7.53 can be written as:

$$I_2(\theta_2) = T_0 I_1(0) \left(\frac{n_2}{n_1}\right)^2 \frac{\cos\theta_2}{\cos\theta_1} \tag{7.55}$$

Since typically $n_2 \ll n_1$, $\cos\theta_1 \simeq 1$ and Eq. 7.55 can be written as:

$$I_2(\theta_2) = T_0 I_1(0) \left(\frac{n_2}{n_1}\right)^2 \cos\theta_2 \tag{7.56}$$

Therefore the light intensity at the output of a planar LED is characterized by a lambertian emission pattern, since it is proportional to $\cos\theta_2$. From Eq. 7.56 it turns out that the intensity is maximum in the direction perpendicular to the semiconductor surface and decreases upon increasing the emission angle θ_2. The polar plot of the lambertian emission is shown in Fig. 7.12 (blue line). The emission pattern can be changed by using non-planar LED surfaces. Indeed, in the case of a hemispherical LED with light-emitting region in the center of the hemisphere, an isotropic emission pattern is obtained (green curve in Fig. 7.12). A highly directional emission can be achieved by using a LED with a parabolic shape as shown by the orange curve in Fig. 7.12.

7.6 Luminous efficiency

An important parameter, which characterizes the performances of a LED, is the *luminous efficiency*, whose definition requires the use of photometric units, which are related to the characteristics of the human eye. The basic photometric unit is the

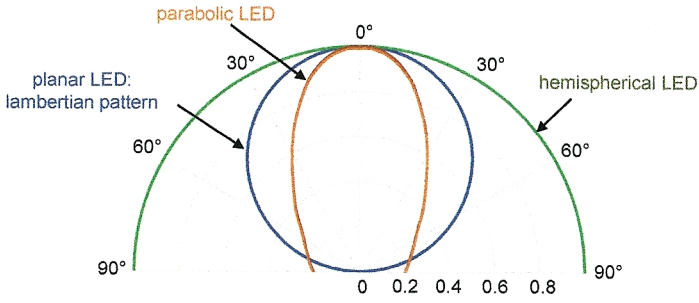

Figure 7.12: Polar plot of the far-field pattern of LEDs with planar (blue curve), hemispherical (green curve) and parabolic (orange curve) surfaces.

luminous flux, Φ_l, which represents the light power emitted by a source spectrally weighted with the human eye's spectral luminous efficiency function $V(\lambda)$, also called luminosity function. The unit of luminous flux is lumen (lm) and it is defined as follows: at 555 nm, where the sensitivity of human eye is maximum (for photopic vision regime), an optical power of 1 W corresponds to a luminous flux of 683 lm. Therefore, in the case of monochromatic radiation at the wavelength λ, the luminous flux, Φ_l, in lumen is given by:

$$\Phi_l = 683\, V(\lambda) P, \tag{7.57}$$

where P is the optical power (in W). The function $V(\lambda)$ is shown in Fig. 7.13 for photopic and scotopic vision regimes. Photopic vision is the vision of human eye under well-lit conditions, while scotopic vision is the vision of the eye under low light conditions. If the light is not monochromatic, the luminous flux can be calculated by

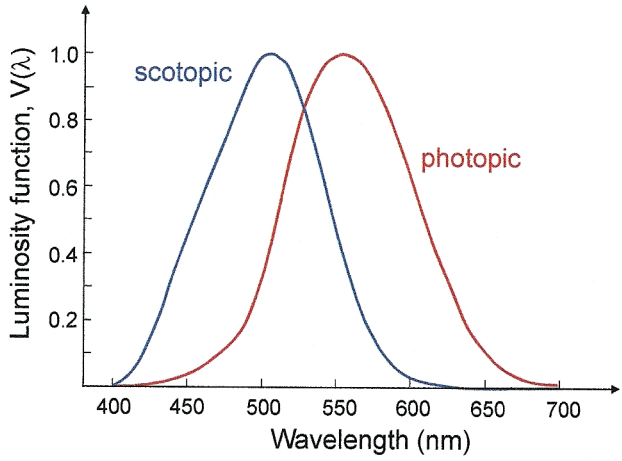

Figure 7.13: Spectral luminous efficiency function, $V(\lambda)$, for photopic (red curve) and scotopic (blue curve) vision regimes.

using the following expression:

$$\Phi_l = 683 \int V(\lambda) P(\lambda)\, d\lambda, \tag{7.58}$$

where $P(\lambda)$ is the power spectral density. The total optical power emitted by a non-monochromatic source is given by:

$$P = \int_{\lambda} P(\lambda)\, d\lambda. \tag{7.59}$$

The luminous efficiency of a LED (measured in units of lm/W) is the ratio between the luminous flux of the LED and the electrical input power:

$$\eta_l = \frac{\Phi_l}{IV} \tag{7.60}$$

where I and V are current and voltage applied to the LED.
Figure 7.14 displays the evolution of LED performances in terms of luminous efficiency. The first red GaAsP LEDs with emission at 650 nm had a luminous efficiency of less than 0.1 lm/W (for comparison an incandescent lamp has an efficiency of $\sim$ 15 lm/W). Current white light LEDs show efficiencies well above 100 lm/W and it is anticipated that by 2020 efficiencies in the 200-lm/W range will be obtained by commercial LEDs.

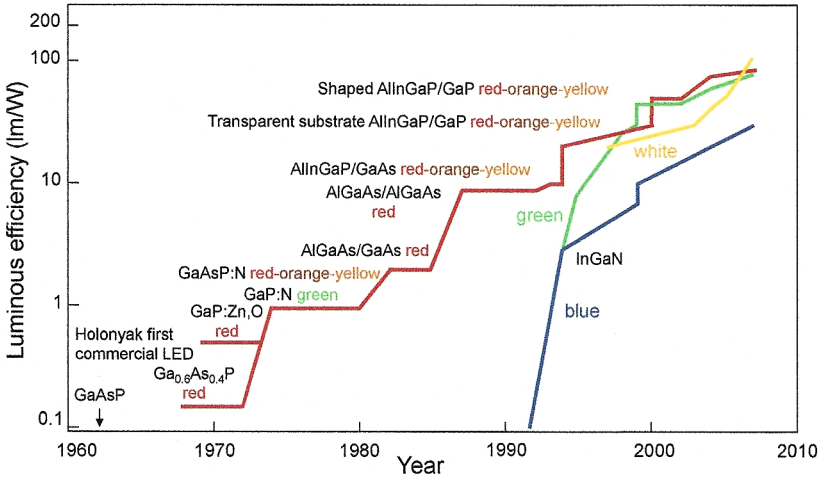

Figure 7.14: Evolution of LED performances in terms of luminous efficiency. From M.G. Craford, Proceeding of the IEEE, **101**, 2170 (2013).

7.7 Blue Light Emitting Diodes

In this last section we well discuss a particularly important class of LEDs: the blue LEDs. Their development has led to the generation of highly efficient white light sources, with important technological applications. Indeed, blue LEDs can be used to excite a phosphor material (like Ce:YAG), which emits in the green and red spectral regions. The combination of the residual blue light with green and red

leads to the generation of white light. Another technique for the generation of white light is based on the combined use of three LEDs of complementary colors (blue, green and red). These white-light LED sources are now progressively replacing the fluorescence lamps for lighting applications, with enormous reduction of electrical energy consumption, since LED-based white light sources require about ten times less energy than typical light bulbs. The invention of efficient blue LEDs was awarded with a Nobel prize in Physics in 2014 to the three Japanese researchers I. Akasaki, H. Amano and S. Nakamura, with the following motivation: "for the invention of efficient blue light-emitting diodes which has enabled bright and energy-saving white light sources."

The material of choice for blue LEDs is GaN (gallium nitride), which belongs to the III-V class, with a wurtzite crystal structure. It has a direct bandgap of 3.4 eV (corresponding to an emission wavelength of 365 nm, in the violet spectral region). GaN can be grown on a substrate of sapphire (Al_2O_3) or SiC: the problem is the large mismatch in the lattice constants ($\sim 16\%$). It can be doped, for example with Si for $n-$type doping and with Mg for $p-$type doping. The growth of high-quality GaN layers on sapphire substrates and the $p-$doping have been the two major problems, which delayed considerably the development of efficient blue LEDs.

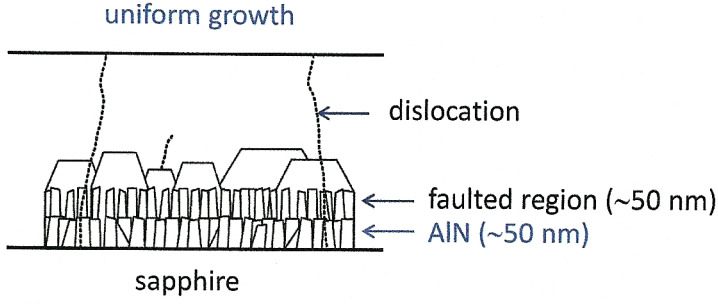

Figure 7.15: Low-temperature buffer layer technology for the growth of high-quality GaN on sapphire (adapted from: K. Hiramatsu *et al.*, J. Crystal Growth **115**, 628 (1991)).

The main idea at the basis of the novel growing technique, which allowed a high-quality deposition of GaN on sapphire was the use of a soft or flexible thin buffer layer between the substrate and the GaN film. This was obtained by developing in 1985 a low-temperature buffer layer technology. First a thin layer ($\sim 30 - 50$ nm) of polycrystalline AlN was deposited on a sapphire substrate at low temperature (500 °C) and then it was heated up to the temperature required for epitaxial growth of GaN single crystal ($\sim$ 1000 °C). The AlN layer is composed of columnar fine crystals, with a diameter of the order of 10 nm: this structure is reached during the heating process, starting from an initial amorphous-like structure (when deposited at low temperature). GaN is subsequently grown on this buffer layer. As shown in Fig. 7.15, the density of dislocations of the GaN crystal is very high in contact to the buffer layer, but then it rapidly decreases after a few-micrometer growth, thus leading to a very high-quality surface. In 1991, by using a similar method, the AlN buffer layer was replaced by a thin GaN buffer layer grown at low temperature.

The second major problem was the $p-$doping of GaN. In 1988 it was observed that the luminescence of a high-quality GaN film with Zn-doping was greatly increased

when it was irradiated by a low-energy electron beam. In the same way the blue luminescence of Mg-doped GaN irradiated by low-energy electrons was significantly increased and the material was converted to a low-resistivity $p-$type crystal. Later it was found that $p-$type GaN can be obtained by thermal annealing above 400 °C in a H_2-free atmosphere and it was shown that Mg is easily passivated by forming Mg-H complexes, so that in order to obtain $p-$doping it was required to activate the Mg acceptor by releasing the hydrogen.

Another very important progress was the growth and $p-$doping of high crystal quality alloys (AlGaN, InGaN) for the development of heterostructure. InGaN is an ideal choice for the active layer in a double heterostructure. Upon changing the amount of In it is possible to tune the wavelength of the emission peak from 400 nm to 445 nm. A challenging aspect is that the introduction of In atoms in a GaN lattice gives rise to a large strain since the radius of In atoms is $\sim 20\%$ bigger than that of Ga atoms.

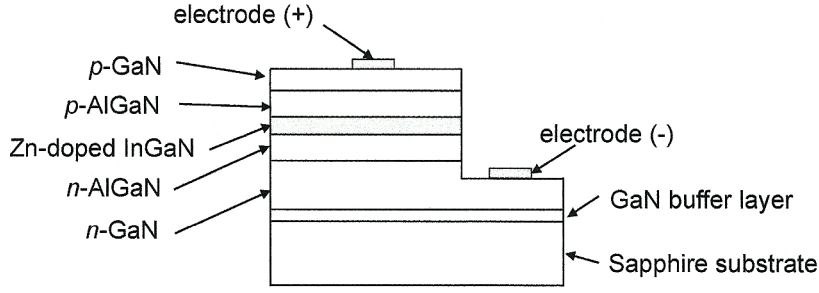

Figure 7.16: Schematic structure of a blue LED with a double heterostructure InGaN/AlGaN.

Figure 7.16 shows the structure of the first commercially available blue LED based on a InGaN/AlGaN double heterostructure with emission wavelength of 450 nm, output power of 1.5 mW and external quantum efficiency (defined in Sect. 7.4) of 2.7%. The structure of this LED is $p-$GaN/$p-$AlGaN/Zn-doped InGaN/$n-$AlGaN/$n-$GaN, with a 45-nm-thick active layer (Zn-doped InGaN). An electron blocking layer was used in the device. The $p-$AlGaN prevents electron overflow from the active layer thus enhancing radiative recombination.

An obvious way to increase the emitted radiation is to increase the input current density: this is particularly effective in high-efficiency devices since less input energy is converted in thermal energy by non-radiative recombination processes. On the other hand, upon increasing the input current density the LED efficiency first increases, reaching a maximum value, and then it decreases if the current density is further increased. This phenomenon is called *efficiency droop* and it is common in all LED devices. For this reason the operational current densities of LEDs are smaller than the maximum current densities that would lead to excessive heating of the structure. A particularly effective physical process leading to efficiency droop is the Auger recombination process, whose efficiency scales with the cube of the carrier concentration (Cn^3) at high injection levels. Due to the high density of carriers in highly efficient LEDs, this term starts to be the dominant (non-radiative) recombination process, thus reducing the internal quantum efficiency of the device.

Blue LEDs, in combination with red and green LEDs allow the development of completely solid-state full-color displays, traffic lights and other specialized lighting applications. Efficient blue LEDs in combination with phosphors to obtain white light are characterized by very long lifetimes ($\sim$ 100000 hours) and are two times more efficient than fluorescent lamps: for this reasons white and UV LEDs are now largely used in TVs, mobile phones and computer displays.

8. Semiconductor Lasers

8.1 Introduction

In Chapter 5 we have seen that a semiconductor far from thermodynamic equilibrium can present a net gain if the Bernard-Duraffourg condition is achieved by proper carrier injection:

$$\mathcal{E}_2 - \mathcal{E}_1 < \mathcal{E}_{Fc} - \mathcal{E}_{Fv}. \tag{8.1}$$

This condition is required in order to generate a carrier inversion and can be obtained, for example, by electrical pumping of a $p-n$ junction composed by degenerate semiconductors, so that both the quasi-Fermi levels, $\mathcal{E}_{Fc}$ and $\mathcal{E}_{Fv}$, fall in the conduction and valence bands, respectively (a very schematic view of a semiconductor laser is shown in Fig. 8.1). Laser action in a semiconductor was first reported in 1962

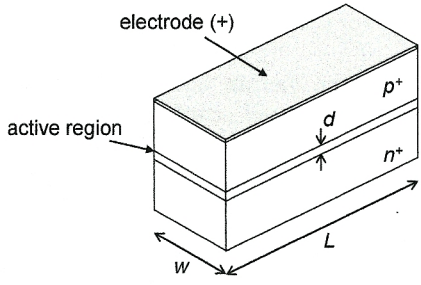

Figure 8.1: Schematic structure of a semiconductor diode laser.

by using $p-n$ homojunctions. As discussed in the previous chapter in the case of light-emitting diodes, homojunctions are not used any more since they present a number of severe drawbacks: they can operate CW only at cryogenic temperatures

($T = 77$ K) due to the high threshold current density at room temperature, of the order of 10^5 A/cm^2. Double-heterostructures present various advantages which lead to a strong reduction of the threshold current density at room temperature ($\sim 10^3$ A/cm^2). Figure 8.2 shows schematically the energy band in the case of a forward-biased double-heterostructure laser. As in the case of LEDs, the active region, with a typical thickness of $0.1 - 0.2$ μm, is placed between the n-type and p-type cladding layers with a larger energy gap. Therefore the generated photons are not absorbed

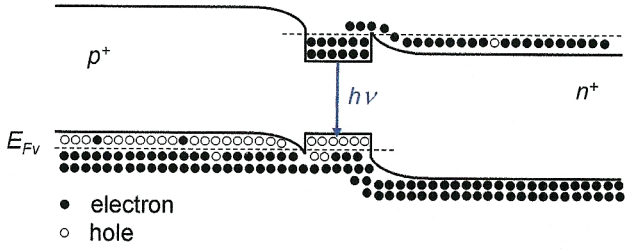

Figure 8.2: Energy band structure of a forward-biased double heterostructure laser.

outside the active region. The energy barriers between the active region and the cladding layers, due to the significantly different energy gaps, confine effectively the injected electrons and holes inside the active layer (*carrier confinement*), thus increasing their concentrations at a given current density. Moreover the refractive index of the active layer is significantly larger than that of the $p-$ and $n-$type cladding layers, thus giving rise to a waveguiding structure, which effectively confines the laser beam in the active layer (*photon confinement*). The number of modes in this waveguide depends on the thickness and on the refractive indexes of the semiconductors. By a proper design a single transverse mode can be obtained. Quantum-well double-heterostructures can be used, as will be discussed in this chapter.

In order to achieve laser action a suitable positive optical feedback is required. In the case of semiconductor lasers Fresnel reflection at the semiconductor-air interfaces are large enough to provide the required feedback for laser action: the parallel end faces are prepared by cleavage along crystal planes. This is due to the high refractive index of semiconductors (e.g., in the case of GaAs $n = 3.6$, thus leading to a reflectance $R = 0.32$) together with the high gain which can be obtained from semiconductors.

In this chapter we will first write the rate equations of a semiconductor laser to deduce the threshold condition for laser action and the output power. Various device structures will be described, which have been introduced to optimize the laser performances. Finally, the main spectral and spatial characteristics of semiconductor laser emission will be discussed.

8.2 Rate equations and threshold conditions for laser action

We will describe the properties of a semiconductor laser by using a set of rate equations. Space-independent rate equations will be written since we assume that the laser oscillates on a single mode, moreover the pumping and mode energy are uniform within the active layer with volume V. Leakage currents will be completely neglected. The mode transverse profile is considered uniform. In the following rate

equations N is the carrier density and ϕ is the photon density:

$$\frac{dN}{dt} = \eta_i \frac{I}{eV} - \frac{N}{\tau} - R_{st} \tag{8.2}$$

$$\frac{d\phi}{dt} = \Gamma R_{st} + \Gamma \beta_{sp} R_{sp} - \frac{\phi}{\tau_c} \tag{8.3}$$

In Eq. 8.2 $\eta_i I/eV$ represents the pumping rate, I is the injected current, η_i is the injection efficiency, giving the fraction of injected current which generates carriers in the active region; N/τ is the recombination rate (radiative and non-radiative), where $1/\tau$ is given by Eq. 7.11 (with $n = N$). R_{st} represents the net stimulated recombination rate, which includes both stimulated absorption and emission.
The first term, ΓR_{st}, of Eq. 8.3 represents the growth rate of the photon density due to stimulated emission. Γ is the *confinement factor*, which will be formally defined below. The simplest definition of the confinement factor is $\Gamma = V/V_p$, where V_p is the cavity volume occupied by photons, which is typically larger than the volume of the active region occupied by electrons, so that the photon density generation rate is $(V/V_p)R_{st} = \Gamma R_{st}$. If we assume a complete overlap between the active region and the optical mode (i.e., $\Gamma = 1$), the growth of the photon density in the cavity can be obtained in terms of the gain per unit length, g, as:

$$\phi(z) = \phi(0)\exp(gz) = \phi(0)\exp(gv_g t), \tag{8.4}$$

where v_g is the group velocity. If we now calculate the time derivative of Eq. 8.4 we obtain the rate of change of the photon density due to stimulated emission, R_{st}:

$$\left(\frac{d\phi}{dt}\right)_{st} = R_{st} = v_g g \phi. \tag{8.5}$$

The term $\Gamma \beta_{sp} R_{sp}$ is related to the photon density generation rate due to spontaneous emission. R_{sp} is the spontaneous recombination rate, $R_{sp} \simeq BN^2$; β_{sp}, called spontaneous emission factor, represents the fraction of the total spontaneous emission, which is coupled into the lasing mode. It can be shown that β_{sp} is the reciprocal of the number of available optical modes in the bandwidth of the spontaneous emission in the case of uniform coupling to all modes. τ_c is the *photon lifetime* in the cavity, which takes into account the loss of photons due to cavity losses and can be written as:

$$\tau_c = \frac{L}{\gamma v_g}, \tag{8.6}$$

where L is the cavity length and γ represents the logarithmic losses:

$$\gamma = \alpha L - \log R, \tag{8.7}$$

where α is the average internal loss and R is the intensity reflection coefficient.

If we assume steady-state conditions ($dN/dt = 0$ and $d\phi/dt = 0$) from Eq. 8.3, assuming that only a very small portion of the spontaneous emission is coupled into the oscillating mode, so that β_{sp} is very small and the term $\Gamma \beta_{sp} R_{sp}$ can be neglected, we have:

$$g_{th}\Gamma = \frac{1}{v_g \tau_c}. \tag{8.8}$$

Equation 8.8 expresses the threshold condition for laser action. Indeed, at threshold the spatially averaged gain must equal the spatially averaged losses:

$$\langle g \rangle_{th} L = \langle \alpha_a \rangle L + \langle \alpha_n \rangle L + \langle \alpha_p \rangle L + \gamma_m, \tag{8.9}$$

where α_a is the scattering loss in the active layer, α_n and α_p are the losses in the $n-$type and $p-$type cladding layers, respectively (in the following we will assume $\alpha_n \simeq \alpha_p = \alpha$) and $\gamma_m = \log(1/R)$ represents the mirror loss. In Eq. 8.9 the averaged values are calculated over the field intensity distribution in the cavity, $|u(x,y,z)|^2$:

$$\langle g \rangle = \frac{\int_a g|u|^2 dV}{\int |u|^2 dV}, \tag{8.10}$$

where the integral at the numerator is calculated over the volume of the active

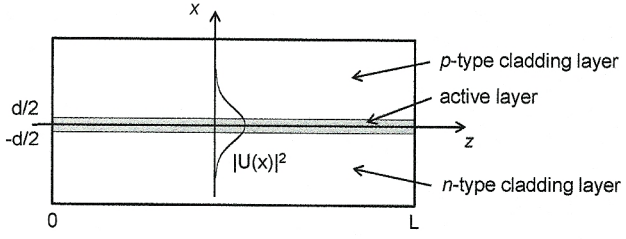

Figure 8.3: Schematic structure of a semiconductor laser showing the field intensity distribution, $|u(x)|^2$, in the cavity.

region and the integral at the denominator is calculated over the whole volume of the laser cavity. $\langle g \rangle$ is called modal gain coefficient and $\langle g \rangle L$ is the modal gain. In the following we will neglect the spatial variation of the field intensity distribution along the longitudinal coordinate z and along the coordinate y parallel to the junction, so that $u = u(x)$ (see Fig. 8.3), where x is the coordinate in the direction perpendicular to the junction. Therefore we can write:

$$\langle g \rangle = \frac{\int_{-d/2}^{d/2} g|u(x)|^2 dx}{\int_{-\infty}^{\infty} |u(x)|^2 dx} = g\Gamma, \tag{8.11}$$

where Γ is the beam confinement factor defined as:

$$\Gamma = \frac{\int_{-d/2}^{d/2} |u(x)|^2 dx}{\int_{-\infty}^{\infty} |u(x)|^2 dx}. \tag{8.12}$$

From the definition of Γ it is clear that the confinement factor gives the fraction of the beam power contained in the active region. From the threshold condition 8.9, by using the definition of Γ we can write:

$$g_{th}\Gamma = \alpha_a \Gamma + \alpha(1-\Gamma) - \frac{1}{L}\log R. \tag{8.13}$$

If we define the average internal loss as:

$$\langle \alpha_i \rangle = \alpha_a \Gamma + \alpha(1-\Gamma), \tag{8.14}$$

the previous equation can be written as:

$$g_{th}\Gamma = \langle \alpha_i \rangle - \frac{1}{L}\log R = \frac{\gamma}{L} = \frac{1}{v_g \tau_c}, \tag{8.15}$$

as given by Eq. 8.8. As it was pointed out in Sect. 5.4, the gain coefficient can be written as a function of the carrier density N as:

$$g = \sigma(N - N_{tr}), \tag{8.16}$$

where σ is the differential gain coefficient and N_{tr} is the carrier density at transparency. We note that the steady-state gain, as deduced from Eq. 8.3, is clamped at the value given by Eq. 8.8, therefore above threshold the carrier density N is clamped at its threshold value:

$$N = N_{th} = \frac{g}{\sigma} + N_{tr} = \frac{\gamma}{\sigma L \Gamma} + N_{tr}. \tag{8.17}$$

It is evident that above threshold the carrier density remains fixed at the critical (threshold) value, while the photon density increases linearly with the injection current I. By using Eq. 8.2 in steady-state conditions it is possible to calculate the current and the current density at threshold. At threshold $\phi = 0$, therefore since $dN/dt = 0$ and $N = N_{th}$, from Eq. 8.2 it is possible to calculate the threshold current (and the threshold current density):

$$I_{th} = \frac{eV}{\eta_i \tau} N_{th}, \qquad J_{th} = \frac{ed}{\eta_i \tau} N_{th}. \tag{8.18}$$

By using Eq. 8.18 and Eq. 8.17 the threshold current density can be written as:

$$J_{th} = \frac{ed}{\eta_i \tau}\left(\frac{\gamma}{\sigma L \Gamma} + N_{tr}\right). \tag{8.19}$$

It is worth to mention that a simple expression for the confinement factor is given by:

$$\Gamma \simeq \frac{V^2}{2 + V^2}, \tag{8.20}$$

where:

$$V = \frac{2\pi d}{\lambda_0}\sqrt{n_a^2 - n_c^2}, \tag{8.21}$$

where n_a and n_c are the refractive indexes of the active medium and of the cladding layers, respectively.

Figure 8.4 shows the evolution of the threshold current density as a function of the thickness, d, of the active layer for a double heterostructure GaAs laser. Upon increasing the active layer thickness, the threshold current density first decreases and reaches a minimum value, then it starts to increase rapidly. This behaviour can be understood on the basis of Eq. 8.19. For sufficiently large values of d ($d > 0.15$ μm), N_{tr} is the dominant term with respect to $\gamma/\sigma L\Gamma$. In this case J_{th} turns out to be proportional to d. When d is very small ($d < 0.1$ μm) also the confinement factor becomes very small. Indeed, by using Eq. 8.20, for very small values of d, Γ becomes proportional to d^2. In this case the first term in the parenthesis of Eq. 8.19 is the dominant one, so that the threshold current density becomes proportional to $1/d$.

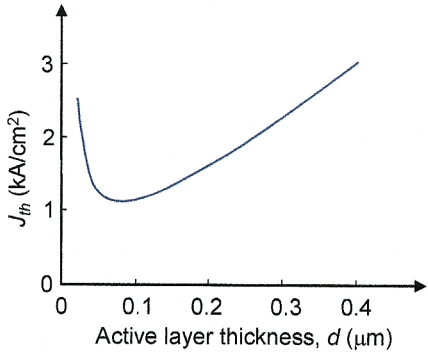

Figure 8.4: Evolution of the threshold current density, J_{th}, as a function of the thickness, d, of the active layer for a 300-μm-long GaAs double heterostructure laser.

Exercise 8.1 Let us consider a GaAs/$Al_{0.3}Ga_{0.7}As$ double heterostructure laser. Calculate the threshold current density assuming the following device parameters: refractive index of GaAs $n_1 = 3.6$, refractive index of $Al_{0.3}Ga_{0.7}As$ $n_2 = 3.4$, thickness of the active medium $d = 0.2$ μm, cleaved an uncoated end faces, absorption coefficient $\alpha = 10$ cm^{-1}, cavity length $L = 250$ μm, differential gain coefficient $\sigma = 1.5 \times 10^{-16}$ cm^2, transparency carrier density $N_{tr} = 2 \times 10^{18}$ cm^{-3}, emission wavelength $\lambda = 850$ nm, injection efficiency $\eta_i = 1$, recombination time constant $\tau = 4$ ns.

We first calculate the confinement factor Γ by using the approximation given in Eq. 8.20, where V is given by (see Eq. 8.21):

$$V = \frac{2\pi d}{\lambda}(n_1^2 - n_2^2) = 1.75$$

$$\Gamma \simeq \frac{V^2}{2+V^2} = 0.60$$

The threshold carrier density is given by Eq. 8.17:

$$N_{th} = \frac{\gamma}{\sigma L \Gamma} + N_{tr}$$

$\gamma = \alpha L - \log R$, where $R = (n_1 - 1)^2/(n_1 + 1)^2 \simeq 0.32$ so that $\gamma = 1.39$.

$$N_{th} = \frac{\gamma}{\sigma L \Gamma} + N_{tr} = (0.61 \times 10^{18} + 2 \times 10^{18})\ \text{cm}^{-3} = 2.61 \times 10^{18}\ \text{cm}^{-3}$$

The threshold current density is given by:

$$J_{th} = \frac{ed}{\eta_i \tau} N_{th} = 2.1 \times 10^3\ \text{A/cm}^2.$$

■

The threshold current has been calculated by assuming a linear dependence of the gain with respect to the carrier density (see Eq. 8.16). In a more general case the dependence of gain vs carrier density can be well approximated by a three-parameter logarithmic expression:

$$g = g_0' \ln \frac{N + N_s}{N_{tr} + N_s} \tag{8.22}$$

Upon increasing N_s in the previous formula the relationship between g and N becomes more linear. For positive gain and when g does not linearly increase with N, a two-parameter approximation gives a very good fitting of the experimental data:

$$g = g_0 \ln \frac{N}{N_{tr}} \tag{8.23}$$

where g_0 is an empirical gain coefficient. The differential gain is given by $dg/dN = g_0/N$. At threshold, by using Eq. 8.15 and Eq. 8.23 it is possible to write:

$$N_{th} = N_{tr} \exp\left(\frac{g_{th}}{g_0}\right) = N_{tr} \exp\left(\frac{\gamma}{\Gamma g_0 L}\right). \tag{8.24}$$

In good laser materials the recombination at threshold is dominated by the spontaneous recombination rate, BN_{th}^2, so that $1/\tau \simeq BN_{th}$ and the threshold current is given by:

$$I_{th} \simeq \frac{eV}{\eta_i} BN_{th}^2. \tag{8.25}$$

By using Eq. 8.24 in 8.25 an expression for the threshold current can be obtained:

$$I_{th} \simeq \frac{eV}{\eta_i} BN_{tr}^2 \exp\left(\frac{2\gamma}{\Gamma g_0 L}\right), \tag{8.26}$$

which clearly shows how I_{th} depends on transparency carrier density, gain, cavity volume and confinement factor. It is evident that to have a small threshold current it is required to reduce the transparency density and to increase the differential gain. In the case of a quantum well laser with a separate confinement waveguide, the small volume V and the small confinement factor Γ may lead to an increase of I_{th}. In this case it can be useful to increase the number of quantum wells (multiple quantum well laser). In this case Eq. 8.26 can still be used by multiplying the volume, V_1, and the confinement factor, Γ_1, of a single well by the number of wells, n_w

$$I_{th} \simeq \frac{en_w V_1}{\eta_i} BN_{tr}^2 \exp\left(\frac{2\gamma}{n_w \Gamma_1 g_0 L}\right). \tag{8.27}$$

The number of wells cannot be increased too much: it is limited by the spatial extension of the optical mode inside the laser cavity.

If the non-radiative recombination events related to Auger processes cannot be neglected, the threshold current increases. In this case:

$$\frac{1}{\tau} \simeq BN_{th} + CN_{th}^2 \tag{8.28}$$

and the threshold current is given by:

$$I_{th} \simeq \frac{eV}{\eta_i}(BN_{th}^2 + CN_{th}^3). \tag{8.29}$$

This is the case, for example, of long-wavelength InGaAsP/InP lasers, where non-radiative recombination events due to Auger processes cannot be neglected. The increase of the threshold current due to the Auger term in Eq. 8.29 is given by:

$$I_{th,nr} = \frac{eV}{\eta_i} C N_{tr}^3 \exp\left(\frac{3\gamma}{\Gamma g_0 L}\right). \tag{8.30}$$

Since this additional term is proportional to the cube of N_{tr}, particular attention has to be paid in these materials to reduce N_{tr}. For InGaAsP at room temperature $C = 2 - 3 \times 10^{-29}$ cm^6/s at 1.3 μm and $C = 7 - 9 \times 10^{-29}$ cm^6/s at 1.55 μm, so that the contribution to the threshold current from the non-radiative processes given by the previous equation becomes the dominant term in Eq. 8.29 for carrier concentrations $N_{th} \simeq 3 \times 10^{18}$ cm^{-3} and $N_{th} \simeq 1.5 \times 10^{18}$ cm^{-3} at 1.3 μm and 1.55 μm, respectively. For this reason particular attention must be paid in these cases to reduce cavity losses and to maintain a large confinement factor. This can be achieved by using strained InGaAs/InGaAsP or InGaAs/InGaAlAs quantum wells on InP. Moreover, the splitting of the valence bands induced by the strain can contribute to reduce the Auger coefficient C.

8.3 Temperature dependence

It is found that in typical double-heterostructure lasers the threshold current changes with temperature as:

$$I_{th}(T) = I_0 \exp\left(\frac{T}{T_0}\right) \tag{8.31}$$

where T is the temperature in kelvin, I_0 is a constant and T_0 is a characteristic temperature (measured in kelvin), which depends on the particular laser diode. T_0 can be used as a measurement of the temperature sensitivity of the diode. Indeed, the larger T_0 the less sensitive is I_{th} to temperature variations, since $\Delta I_{th}/I_{th} = \Delta T/T_0$.

In Eq. 8.26 the transparency density N_{tr}, the gain g_0 and the internal losses (γ factor) typically depend on temperature. In particular, N_{tr} increases with temperature while g_0 decreases with temperature, since at high temperature the injected carriers can occupy a larger energy interval. The characteristic temperature T_0 is measured experimentally and depends on a large number of factors, including the physical properties of the semiconductors used in the laser diode and the particular laser structure. For example, in the case of GaAs/AlGaAs double heterostructure laser $T_0 \geq 120$ K around room temperature. In the case of GaAs/AlGaAs quantum well lasers, T_0 is in the range 150-180 K, while for strained InGaAs/AlGaAs quantum well lasers $T_0 \geq 200$ K. The increase of the characteristic temperature in quantum well lasers is another advantage of the quantum well devices and it is mainly a result of the weaker dependence of the differential gain on temperature. In the case of InGaAsP/InP double heterostructure and quantum well lasers emitting between 1.3 and 1.55 μm, T_0 is in the range of 50-70 K at room temperature. These quite low values are mainly related to Auger recombination and to leakage current and limit the performances of these lasers under high-temperature operation (at 400 K T_0 is of the order of 30 K).

8.4 Output power

In this section a general expression for the output power of a laser diode as a function of the driving current will be calculated. Under steady-state conditions, above threshold, Eq. 8.2 can be re-written as:

$$\eta_i \frac{I}{eV} - \frac{N_{th}}{\tau} - R_{st} = 0. \tag{8.32}$$

By using Eq. 8.5 and Eq. 8.18, the previous equation allows one to calculate the photon density ϕ corresponding to the injection current I:

$$\phi = \frac{\eta_i (I - I_{th})}{eV v_g g_{th}}, \tag{8.33}$$

where g_{th} is given by Eq. 8.15. The optical energy stored in the laser cavity can be calculated as follows:

$$\mathcal{E}_{st} = \phi h\nu V_p \tag{8.34}$$

where $h\nu$ is the energy of the emitted photons. The photon loss rate due to the mirror transmission is given by $-(v_g/L)\log R$, so that the output power is:

$$P = \mathcal{E}_{st}\left(-\frac{v_g}{L}\log R\right) = \eta_i \frac{h\nu}{e}\frac{-\log R}{\alpha L - \log R}(I - I_{th}). \tag{8.35}$$

It is useful to define the external quantum efficiency, η_{ext}, as the ration between the increase in emitted photons and the corresponding increase in injected carriers:

$$\eta_{ext} = \frac{d(P/h\nu)}{d(I/e)} = \frac{e}{h\nu}\frac{dP}{dI} = \eta_i \frac{-\log R}{\alpha L - \log R}. \tag{8.36}$$

By using this definition, the output power can be written as (when $I > I_{th}$):

$$P = \eta_{ext}\frac{h\nu}{e}(I - I_{th}). \tag{8.37}$$

The previous equation gives to total power out of both end faces of the diode. If one mirror has zero transmission, all power is emitted out of the other mirror.

Another useful parameter is the *slope efficiency* of the laser, defined as:

$$\eta_s = \frac{dP}{V_{el} dI}, \tag{8.38}$$

where V_{el} is the applied voltage. From Eqs. 8.36 and 8.38 we see that the relation between the external quantum efficiency and the slope efficiency is:

$$\eta_{ext} = \eta_s \frac{eV_{el}}{h\nu}. \tag{8.39}$$

From Eq. 8.37 it is evident that the $P - I$ curve of a semiconductor laser diode above threshold is a straight line with a slope defined by η_s. Below the threshold current the output power is due to spontaneous emission and depends on the portion of spontaneous emission in the lasing mode. Figure 8.5 shows schematically the evolution of the output power and of the carrier density as a function of the injected current.

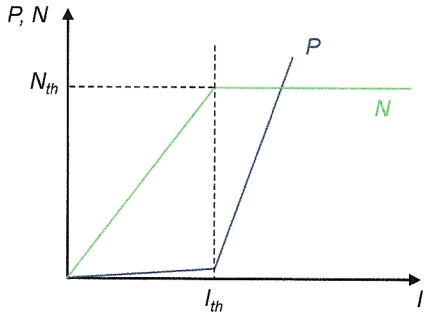

Figure 8.5: Qualitative evolution of the output power, P, and of the carrier density, N, as a function of the driving current.

Below threshold the stimulated emission term, ΓR_{st}, can be neglected in Eq. 8.3, so that, in steady-state conditions, we can write:

$$\phi = \Gamma \beta_{sp} R_{sp} \tau_c. \tag{8.40}$$

If the stimulated emission rate, R_{st}, is neglected in Eq. 8.2, in steady-state conditions we have:

$$\eta_i \frac{I}{eV} = \frac{N}{\tau}, \tag{8.41}$$

which can be rewritten in terms of R_{sp}:

$$\eta_i \frac{I}{eV} = \frac{R_{sp}}{\eta_{int}}. \tag{8.42}$$

where $\eta_{int} = \tau/\tau_{rad}$ is the internal quantum efficiency. The output power can be calculated as before:

$$\begin{aligned} P &= \phi h\nu V_p \left(-\frac{v_g}{L}\log R\right) = \frac{V}{V_p}\beta_{sp}\eta_{int}\eta_i \frac{I}{eV}\frac{L}{v_g(\alpha L - \log R)} h\nu V_p \left(-\frac{v_g}{L}\log R\right) \\ &= \eta_{int}\eta_i \frac{h\nu}{e}\frac{-\log R}{\alpha L - \log R}\beta_{sp} I \end{aligned} \tag{8.43}$$

At threshold the spontaneous emission rate clamps to the value N_{th}/τ_{rad}.

8.5 Quantum well lasers

Quantum wells can be used as active materials in diode lasers to take advantage of the good optical properties of quantum wells compared to the corresponding bulk materials: in particular the typically larger differential gain and weaker dependence of this gain on temperature and the possibility to use strained layers to improve the optical characteristics and the bandgap tunability. Due to the very small thickness of the active material the corresponding confinement factor Γ is very small, thus giving rise to a large threshold current. In order to increase the confinement factor separate confinement heterostructures (SCHs) are used, with various configurations, as shown in Fig. 8.6, which displays schematically the conduction band structure of a simple SCH and of a graded-index SCH. At the center of both structures there is a

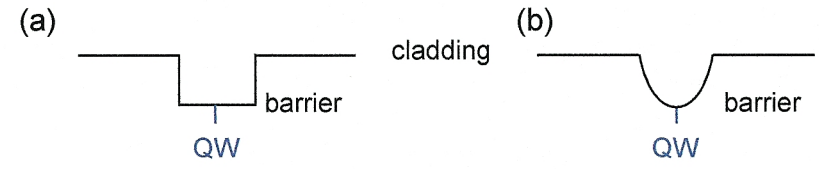

Figure 8.6: Qualitative plot of the conduction energy band of (a) a step-index separate confinement quantum well heterostructure and of (b) a graded-index separate confinement quantum well heterostructure.

thin quantum well, with a thickness of a few nanometers. The beam confinement is achieved by the barrier layers (with a thickness of the order of 0.1 μm), characterized by a higher refractive index compared to the cladding layers. The contribution of the quantum well to beam confinement is negligible, while the carriers are confined by the quantum well. In the case of the graded-index SCH (called GRIN-SCH, from GRaded-INdex SCH) the refractive index of the inner confinement structure is smoothly changed from a maximum value at the center of the structure to the value in the cladding region. For example, in a GaAs/AlGaAs quantum well laser, where GaAs is the well material, the composition of the inner barrier $Al_xGa_{1-x}As$ changes from, e.g., $x = 0.3$ at the center to $x = 0.7$ at the interface with the two cladding layers formed by $Al_{0.7}Ga_{0.3}As$.

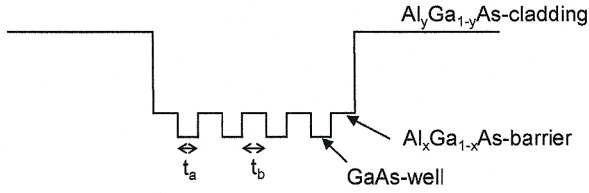

Figure 8.7: Qualitative plot of the conduction energy band of a multiple quantum well laser.

The active material can be a single quantum well or a multiple quantum well, which consists of a few alternating thin layers of wide and narrow energy gap semiconductors, as schematically shown in Fig. 8.7. By using a separate confinement structure it is possible to significantly increase the confinement factor. A very good and simple approximation for the confinement parameter can be obtained as follows. The structure shown in Fig. 8.7 can be analysed considering a three-region waveguide composed by the same cladding layers and by a central layer of thickness:

$$t = N_a t_a + N_b t_b, \tag{8.44}$$

where t_a and t_b are the thickness of the active (well) and barrier layers, respectively (see Fig. 8.7); N_a and N_b are the number of well and barrier layers, respectively. The central region is characterized by an average refractive index given by:

$$n = \frac{N_a t_a n_a + N_b t_b n_b}{t}, \tag{8.45}$$

where n_a and n_b are the refractive indexes of the active and barrier layers, respectively. This equivalent three-region waveguide is characterized by a confinement factor Γ

given by Eq. 8.20, where:

$$V = \frac{2\pi t}{\lambda_0}\sqrt{n^2 - n_c^2}. \tag{8.46}$$

The confinement factor of the real multiple quantum well structure must take into account the fraction of the modal energy in the active regions and can be calculated as:

$$\Gamma_{MQW} = \Gamma\frac{N_a t_a}{t}. \tag{8.47}$$

Note that this simple three-region waveguide model is a fairly accurate model when the waveguide supports only the fundamental mode (the approximation is still relatively good also when higher-order modes can propagate).

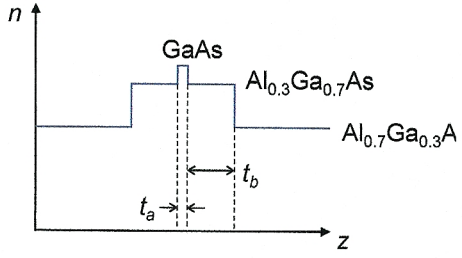

Figure 8.8: Variation of the refractive index of the quantum well diode laser with separate confinement heterostructure analysed in Ex. 8.2.

Exercise 8.2 Consider a quantum well diode laser (with a single well) with a separate confinement heterostructure with the following characteristics. Active layer: GaAs, thickness $t_a = 10$ nm, refractive index $n_a = 3.6$; inner barrier layers: $Al_{0.3}Ga_{0.7}As$, thickness $t_b = 0.1$ μm, refractive index $n_b = 3.4$; cladding layer: $Al_{0.7}Ga_{0.3}As$, refractive index $n_c = 3.18$; emission wavelength $\lambda_0 = 850$ nm. Calculate the confinement factor.

The variation of the refractive index in the laser structure is shown in Fig. 8.8. The three-region equivalent waveguide has a central layer with thickness:

$$t = N_a t_a + N_b t_b = 0.21\ \mu\text{m},$$

where $N_a = 1$ and $N_b = 2$. The average refractive index in this layer is given by Eq. 8.45:

$$n = \frac{N_a t_a n_a + N_b t_b n_b}{t} = 3.41.$$

The V-number is:

$$V = \frac{2\pi t}{\lambda_0}\sqrt{n^2 - n_c^2} = 1.91,$$

the confinement factor of the three-region waveguide is:

$$\Gamma = \frac{V^2}{2+V^2} = 0.65$$

and the corresponding confinement factor of the quantum well structure is:

$$\Gamma_{QW} = \Gamma \frac{t_a}{t} = 0.031.$$

Exercise 8.3 Use the confinement factor obtained in the previous exercise to calculate the threshold current density assuming: $N_{tr} = 2 \times 10^{18}$ cm^{-3}, differential gain $\sigma = 6 \times 10^{-16}$ cm^2, absorption coefficient $\alpha = 10$ cm^{-1}, cavity length $L = 250$ μm, injection efficiency $\eta_i = 1$, recombination time constant $\tau = 4$ ns, mirror reflectivity $R = 0.32$.

As we have calculated in Ex. 8.1 $\gamma = 1.39$. The threshold carrier density is given by:

$$N_{th} = \frac{\gamma}{\sigma L \Gamma} + N_{tr} = (3 \times 10^{18} + 2 \times 10^{18})\ \text{cm}^{-3} = 5 \times 10^{18}\ \text{cm}^{-3}$$

so that the threshold current density is given by:

$$J_{th} = \frac{e t_a}{\eta_i \tau} N_{th} = 200\ \text{A/cm}^2.$$

8.6 Laser structures

The simple laser structure shown in Fig. 8.1 cannot be used, also in the case of double heterostructure configuration, due to the high driving current and the poor quality of the output beam. Lateral confinement of carriers and photons is required in order to achieve stable laser operation and good quality output beams. A lateral confinement of the injected carriers is required to avoid leakage currents, which may bypass the active region, thus strongly reducing the laser efficiency. Several laser configurations have been developed, which can be grouped in two main classes: gain-guided structures and index-guided structures. The *gain-guided* is the simplest structure: Fig. 8.9(a) shows a scheme of an edge-emitting gain guided semiconductor laser. By adding an oxide layer the current injected from the positive electrode is forced to flow in a stripe with narrow width ($w \simeq 5$ μm) thus strongly reducing the required current flow ($I = JLw$). Also the width of the pumped active layer is roughly equal to w thus leading to a (relatively weak) lateral waveguiding, which weakly confines the beam transverse dimension in the direction parallel to the junction. The structure is called gain-guided since the optical beam is mainly guided by the variation of the gain. If the stripe width is < 10 μm the laser mode is confined to the fundamental transverse mode in the direction parallel to the junction, while in the direction perpendicular to the junction the optical beam is confined to the fundamental mode by the double heterostructure or by the separate confinement structure. A major disadvantages of the gain-guided structure is that the un-pumped

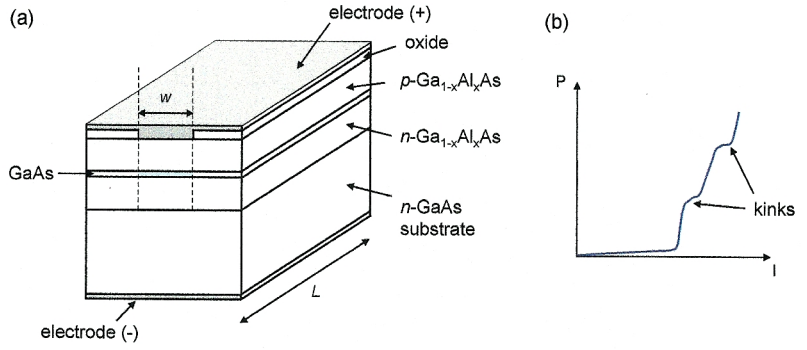

Figure 8.9: (a) Typical structure of gain-guided laser. (b) Power-current curve of a gain-guided laser with kinks associated to unstable lateral waveguiding

portions of the active layer are strongly absorbing thus introducing losses. Moreover, since the lateral waveguiding effect is rather weak, small variations of the refractive index, due for example to variations of the temperature or of the injected carriers, may lead to unstable operation. Threshold current is typically rather high, of the order of 50-100 mA and the power-current curves generally present kinks, a shown in Fig. 8.9(b), associated to unstable lateral waveguiding. The optical spectrum of the output beam is multimodal due to the enhanced spontaneous emission.

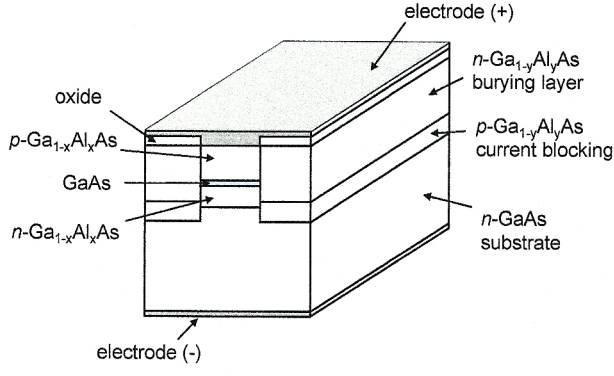

Figure 8.10: Typical structure of a buried heterostructure index-guided laser.

A typical structure of a (strongly) index-guided edge-emitting laser is shown in Fig. 8.10. In this case a small stripe of the active material is embedded into semiconductor layers with larger bandgap (and lower refractive index). Due to reduced lateral width w of the active layer, lateral single-mode operation is possible (w must be smaller than a critical value for higher-order lateral modes). Figure 8.10 refers to a buried-heterostructure GaAs/AlGaAs index-guided laser. The $n-$type burying layer and the $p-$type current blocking layers form a reverse-biased diode, which does not allow the current to bypass the active region. By using strong index-guided laser structures very stable laser performances can be achieved, with lateral

single-mode operation and low threshold currents (<10 mA). For these reasons they are largely used in commercial devices, despite the rather complex fabrication procedures required for the construction.

8.7 Spectral and spatial characteristics of diode laser emission

The output spectrum of a diode laser depends on the structure of the optical resonator and on the optical gain curve of the active medium. In a Fabry-Perot (FP) cavity the length L determines the separation between the longitudinal modes. In a typical semiconductor laser, while below threshold a large number of FP cavity resonances are visible across a wide spectrum, above threshold only a few modes close to the peak of the gain curve oscillate, as shown in Fig. 8.11 in the case of a gain-guided laser emitting at $\sim$ 779 nm. The frequency separation between adjacent modes is given by:

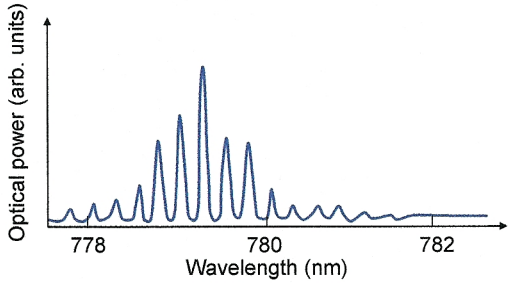

Figure 8.11: Output spectrum of a multi-longitudinal mode gain-guided laser.

$$\Delta\nu = \frac{c}{2Ln_g} \tag{8.48}$$

where n_g is the group index, which takes into account the dispersion of the waveguide:

$$n_g = n_{eff} + \nu\frac{dn_{eff}}{d\nu} = n_{eff} - \lambda\frac{dn_{eff}}{d\lambda} \tag{8.49}$$

n_{eff} is the effective refractive index experienced by an optical mode. For a typical FP laser diode, assuming $n_g = 3.5$ and a cavity length $L = 300$ μm, the mode separation is $\Delta\nu \simeq 140$ GHz, which is rather small in comparison to the typical linewidth of the gain curve of a semiconductor, which can be of the order of a few THz, so that longitudinal multimode operation is achieved above threshold in typical FP diode lasers. A possible way to obtain single-mode operation is to reduce the cavity length L so that the mode spacing exceeds the width of the gain curve: this requires cavity lengths of the order of a few micrometers. These very short cavities can be generated along the epitaxial deposition axis and the corresponding devices are called vertical cavity surface emitting lasers (VCSELs) (see Ch.10). Another technique used for single-mode laser operation is based on the use of distributed Bragg reflector (DBR) lasers or of distributed feedback (DFB) lasers, which will be analysed in Chapt. 9. We note that single-mode laser operation is particularly important for optical-fibre communications, due to the chromatic dispersion of an optical fibre.

In the case of the edge-emitting lasers discussed in the previous section, the output beam is typically elliptical. Due to small beam size in the direction perpendicular to the junction, the output beam is always diffraction limited in the plane perpendicular to the junction. If the stripe width is shorter than ~ 10 μm the beam is diffraction-limited also in the plane parallel to the junction. If a Gaussian beam profile is assumed in both transverse directions, the beam divergences $\theta_{\parallel}$ and $\theta_{\perp}$ in the planes parallel and perpendicular to the junction, respectively, are given by $\theta_{\parallel} = 2\lambda/\pi d_{\parallel}$ and $\theta_{\perp} = 2\lambda/\pi d_{\perp}$ where $d_{\parallel}$ and $d_{\perp}$ are the beam dimensions in the two directions, defined as the full-width between $1/e$ points of the electric field amplitude. Therefore, assuming an elliptical beam on the exit face of a laser with dimensions 1 μm $\times$ 5 μm, the divergence is larger in the direction perpendicular to the junction, so that the beam ellipticity rotates by $90°$ after a few tens of micrometers at the output of the laser cavity. When a circular beam is required, suitable optical systems can be used to compensate for the beam astigmatism.

9. Distributed Feedback Lasers

9.1 Basic concepts

The structure of a distributed feedback (DFB) laser is schematically shown in Fig. 9.1 and it is rather different in comparison to the typical laser schemes that we have seen in the previous chapter. The laser consists of an active medium in which a periodic thickness variation of the cladding layers forms a grating, which provides the feedback required for laser operation. The laser cavity mirrors are distributed along the length of the cavity. The example of Fig. 9.1 refers to a DFB laser oscillating at 1.55 μm. The thickness of the upper InGaAsP confinement layer is periodically modulated, with a period Λ. Due to the thickness modulation the refractive index

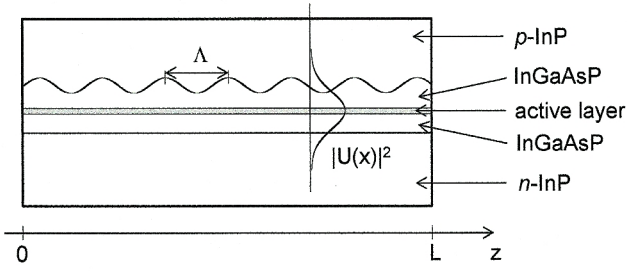

Figure 9.1: Schematic structure of a DFB laser with uniform grating. $|U(x)|^2$ is the transverse distribution of the beam intensity inside the cavity.

experienced by the oscillating mode is periodically modulated, since the refractive index of the InGaAsP confinement layer is larger than that of the InP p- and n-type layers. The effective refractive index can be calculated as $n(z) = \langle n(x,z) \rangle_x$, where the average is calculated over the x-coordinate, perpendicular to the longitudinal propagation z-axis, with a weight determined by the transverse distribution of the beam intensity inside the cavity, $|U(x)|^2$. In the following we will assume a sinusoidal

modulation:

$$n(z) = n_0 + \delta n \cos\left(\frac{2\pi}{\Lambda} z + \phi\right), \tag{9.1}$$

where $\delta n \ll n_0$. The electric field in the forward (positive z) and backward (negative z) directions are effectively coupled when Bragg condition is satisfied:

$$\frac{\lambda_0}{4\bar{n}} = \frac{\Lambda}{2}, \tag{9.2}$$

where λ_0 is the free-space wavelength of the radiation and $\bar{n}$ is the effective refractive index: $\bar{n} = \langle n(z) \rangle$, where the average is calculated along the z coordinate. Note that when the modulation of the refractive index is given by a periodic square-wave function of period Λ, the structure shown in Fig. 9.1 is equivalent to a sequence of alternating layers with high and low refractive index. The Bragg condition for constructive interference of all the waves reflected at each layer interface occurs when the layer thickness is equal to a quarter of the wavelength: $d_1 n_1 = d_2 n_2 = \lambda_0/4$, with $d_1 + d_2 = \Lambda$. Equation 9.2 shows that the grating period Λ determines the wavelength of the mode which can oscillate in the cavity.

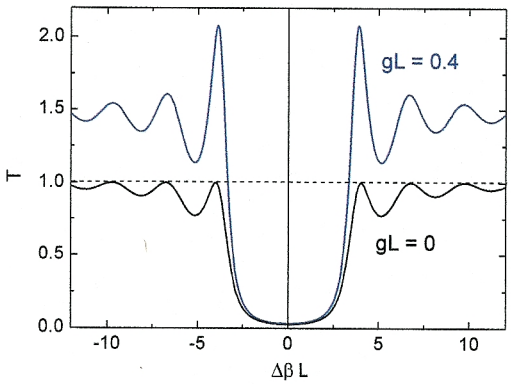

Figure 9.2: Intensity transmittance T of a uniform grating as a function of the detuning $\Delta\beta L$ calculated for two different values of the effective gain gL.

Before proceeding with an analytical treatment of a DFB laser by using the method of the coupled-mode equations, we will anticipate a few general results. We will first assume that the field reflectivities at the two laser end faces are zero ($r_1 = r_2 = 0$). The electric field travelling in the forward direction will be indicated as $E_f(z)$. Moreover, the propagation constant at the Bragg condition is given by:

$$\beta_0 = k_0 \bar{n} = \frac{2\pi}{\lambda_0} \bar{n} = \frac{\pi}{\Lambda}. \tag{9.3}$$

The analytical treatment will show that the intensity transmittance, T, for the forward beam, defined as:

$$T = \frac{|E_f(L)|^2}{|E_f(0)|^2}, \tag{9.4}$$

where L is the cavity length, as a function of the detuning from the Bragg condition, $(\beta - \beta_0)L$, shows maxima, whose values increase upon increasing the gain. The threshold gain corresponds to the gain value required to obtain a strong increase of one (or two) of these maxima. The situation is illustrated in Fig. 9.2, which shows the transmittance T as a function of detuning for two values of the effective gain gL. This figure shows that in the case of a uniform grating (i.e., a grating with constant periodicity along the cavity length) the transmittance is minimum at resonance ($\beta = \beta_0$) and presents various maxima symmetrically located with respect to the Bragg condition. In this case, just above threshold, two modes oscillate at two different frequencies, which correspond to the two maxima closest to $\beta = \beta_0$. The region between these two peaks defines the so-called *stop-band* of the device. Upon increasing the gain, higher-order modes will start to grow, corresponding to the other transmittance maxima.

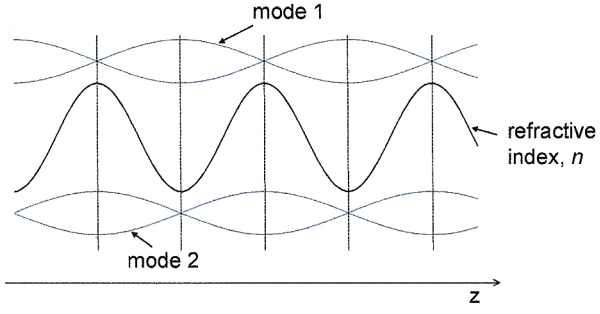

Figure 9.3: Schematic representation of the refractive index n and of the standing-wave patterns of the two lowest-order modes along the cavity of a DFB laser with a uniform grating.

The existence of two oscillating modes can be intuitively understood with the help of Fig. 9.3, which shows the modulation of the refractive index along the cavity together with the standing-wave patterns of the two lowest-order oscillating modes. The effective refractive index experienced by mode 1, $\bar{n}_1$, is smaller than that of mode 2, $\bar{n}_2$, since the anti-nodes of mode 1 are located in correspondence of the minima of the refractive index variation along the cavity, while the anti-nodes of mode 2 are placed in correspondence of the maxima of n. The effective refractive index is calculated as a weighted average along the z-coordinate, where the weight function is the standing wave pattern of the mode. Since $\bar{n}_1 < n_0 < \bar{n}_2$, fre free-space wavelengths of mode 1 and 2, given by $\lambda_{01,2} = 2\bar{n}_{1,2}\Lambda$, are $\lambda_{01} < \lambda_0 < \lambda_{02}$ (inside the cavity the two modes have the same wavelength $\lambda = 2\Lambda$).

In general, this double-mode operation is not advantageous for applications; various techniques have been introduced to achieve single-mode operation. For example it is possible to use an asymmetric structure, characterized by two different mirror reflectivities r_1 and r_2. The best solution is represented by the use of a periodic modulation of the thickness of the cladding layer, which undergoes a phase shift of $\Lambda/2$ at the center of the structure. Since at the Bragg condition (Eq. 9.2) $\Lambda/2 = \lambda_0/4\bar{n}$, this structure is called $\lambda/4$-*shifted DFB laser*. Figure 9.4 shows the transmittance of a $\lambda/4$-shifted DFB device as a function of the detuning, for two values of the effective gain. A maximum is present at the Bragg resonance: just above the threshold gain, a single mode will oscillate with $\beta = \beta_0$. Upon increasing

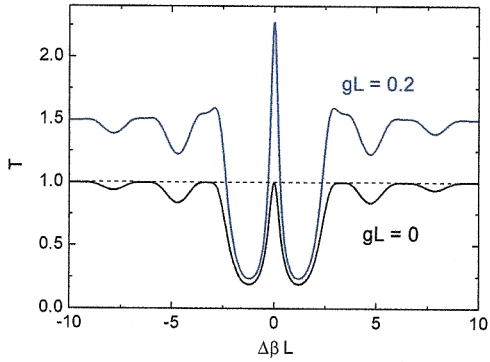

Figure 9.4: Intensity transmittance T of a $\lambda/4$-shifted grating as a function of the detuning, $\Delta\beta L$, calculated for two different values of the effective gain gL.

the gain above the threshold value for single-mode operation, higher-order modes, symmetrically located with respect to the Bragg frequency, will start to grow. In this case the mode selectivity is higher than in the case of DFB lasers with uniform grating, i.e., the relative gain increase with respect to the single-mode condition in order to have oscillation of the higher-order modes is higher than in the case of uniform grating. The existence of a single maximum at the Bragg frequency can be intuitively understood following the same arguments used in the previous case. The modulation of the refractive index along the cavity and the standing-wave pattern of the lowest order mode is shown in Fig. 9.5. The effective refractive index experienced by the mode is n_0 and the free-space wavelength is $\lambda_0 = \lambda_B = 2\Lambda n_0$.

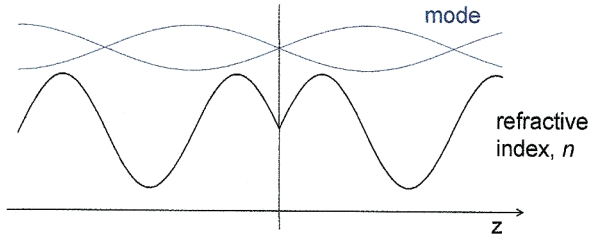

Figure 9.5: Schematic representation of the refractive index, n, and of the standing-wave pattern of the lowest-order mode along the cavity of a DFB laser with a $\lambda/4$-shifted grating.

9.2 Coupled-mode theory

We will first consider a passive structure ($g = 0$). The refractive index modulation is given by:

$$n(z) = n_0 + \delta n\, h(z) \cos\left(\frac{2\pi}{\Lambda} z + \phi(z)\right), \tag{9.5}$$

where $\delta n \ll n_0$ and the functions $h(z)$ (with $0 < h(z) < 1$) and $\phi(z)$ have been introduced to analyse a general situation, where both the amplitude and phase of the modulation are functions of the propagation coordinate z. The general wave equation is:

$$\nabla^2 \widetilde{\mathbf{E}} - \frac{n^2}{c^2} \frac{\partial^2 \widetilde{\mathbf{E}}}{\partial t^2} = 0. \tag{9.6}$$

We assume that the electric field distribution in the xy plane, perpendicular to the longitudinal coordinate z, is given by the fundamental mode of the waveguide:

$$\widetilde{\mathbf{E}}(x, y, z, t) = F(x, y) E(z, t) \mathbf{e}, \tag{9.7}$$

where $\mathbf{e}$ is the unit vector in the direction of the field polarization. From Eqs. 9.6 and 9.7 we obtain the following wave equation:

$$\frac{\partial^2 E}{\partial z^2} - \frac{n^2(z)}{c^2} \frac{\partial^2 E}{\partial t^2} = 0. \tag{9.8}$$

We can write the general solution as the sum of two counter-propagating modes:

$$E = A(z, t) \exp[i(\beta_0 z - \omega_0 t)] + B(z, t) \exp[-i(\beta_0 z + \omega_0 t)] + \text{c.c.} \tag{9.9}$$

where $A(z, t)$ and $B(z, t)$ are the amplitudes of the forward and backward fields, respectively. Since the coupling is weak, we will assume that the field amplitudes vary slowly, both in space and in time. Therefore we have:

$$\left|\frac{\partial A}{\partial z}\right| \Lambda \ll |A| \qquad \left|\frac{\partial^2 A}{\partial z^2}\right| \ll \frac{1}{\Lambda} \left|\frac{\partial A}{\partial z}\right| = \frac{\beta_0}{\pi} \left|\frac{\partial A}{\partial z}\right| \tag{9.10}$$

$$\left|\frac{\partial A}{\partial t}\right| \frac{2\pi}{\omega_0} \ll |A| \qquad \left|\frac{\partial^2 A}{\partial t^2}\right| \ll \omega_0 \left|\frac{\partial A}{\partial t}\right| \tag{9.11}$$

The same expressions hold for the amplitude $B(z, t)$ of the backward wave. By using the slowly-varying envelope approximation, the spatial and temporal derivatives of $E(z, t)$ in Eq. 9.8 can be written as:

$$\begin{aligned} \frac{\partial^2 E}{\partial z^2} &= \left(2i\beta_0 \frac{\partial A}{\partial z} - \beta_0^2 A\right) e^{i(\beta_0 z - \omega_0 t)} + \\ &- \left(2i\beta_0 \frac{\partial B}{\partial z} + \beta_0^2 B\right) e^{-i(\beta_0 z + \omega_0 t)} + \text{c.c.} \end{aligned} \tag{9.12}$$

$$\begin{aligned} \frac{\partial^2 E}{\partial t^2} &= -\left(2i\omega_0 \frac{\partial A}{\partial t} + \omega_0^2 A\right) e^{i(\beta_0 z - \omega_0 t)} + \\ &- \left(2i\omega_0 \frac{\partial B}{\partial t} + \omega_0^2 B\right) e^{-i(\beta_0 z + \omega_0 t)} + \text{c.c.} \end{aligned} \tag{9.13}$$

Moreover $n^2(z)$ can be written as follows:

$$n^2(z) \simeq n_0^2 + 2n_0 \delta n \, h(z) \cos\left(\frac{2\pi}{\Lambda} z + \phi(z)\right) = n_0^2 + 2n_0 \delta n \, h(z) \cos(2\beta_0 z + \phi) \tag{9.14}$$

or, in a similar way:

$$n^2(z) \simeq n_0^2 + n_0 \delta n \, h(z)[\exp(2i\beta_0 z)e^{i\phi} + \exp(-2i\beta_0 z)e^{-i\phi}] \tag{9.15}$$

By substituting Eqs. 9.12, 9.13 and 9.15 in the wave equation 9.8 and equating the terms containing $\exp[i(\beta_0 z - \omega_0 t)]$ we obtain:

$$\begin{aligned} &2i\beta_0 \frac{\partial A}{\partial z} - \beta_0^2 A + \frac{n_0^2}{c^2}\left(2i\omega_0 \frac{\partial A}{\partial t} + \omega_0^2 A\right) + \\ &+ \frac{n_0 \delta n \, h(z)}{c^2} e^{i\phi}\left(2i\omega_0 \frac{\partial B}{\partial t} + \omega_0^2 B\right) = 0 \end{aligned} \tag{9.16}$$

Since $\beta_0 = \omega_0 n_0 / c$

$$-\beta_0^2 A + \frac{n_0^2 \omega_0^2}{c^2} A = 0. \tag{9.17}$$

If we neglect the term $\partial B/\partial t$ with respect to $\omega_0 B$ from Eq. 9.16 we get:

$$\frac{\partial A}{\partial z} + \frac{n_0}{c}\frac{\partial A}{\partial t} = i\frac{\delta n}{2n_0}\beta_0 h(z) e^{i\phi(z)} B \tag{9.18}$$

In the same way, by substituting Eqs. 9.12, 9.13 and 9.15 in the wave equation 9.8 and equating the terms containing $\exp[-i(\beta_0 z + \omega_0 t)]$ we can write:

$$\frac{\partial B}{\partial z} - \frac{n_0}{c}\frac{\partial B}{\partial t} = -i\frac{\delta n}{2n_0}\beta_0 h(z) e^{-i\phi(z)} A \tag{9.19}$$

The last two equations show that the forward and backward waves are coupled by the modulation of the refractive index. A coupling parameter, $q(z)$, is usually introduced:

$$q(z) = \frac{\delta n}{2n_0}\beta_0 h(z) e^{i\phi(z)} \tag{9.20}$$

so that the coupled-mode equations can be re-written as follows:

$$\begin{cases} \dfrac{\partial A}{\partial z} + \dfrac{n_0}{c}\dfrac{\partial A}{\partial t} = iq(z)B \\ \\ \dfrac{\partial B}{\partial z} - \dfrac{n_0}{c}\dfrac{\partial B}{\partial t} = -iq^*(z)A \end{cases} \tag{9.21}$$

So far we have considered a passive structure; the effect of gain can be simply introduced in Eqs. 9.21:

$$\begin{cases} \dfrac{\partial A}{\partial z} + \dfrac{n_0}{c}\dfrac{\partial A}{\partial t} = iq(z)B + \dfrac{g}{2}A \\ \\ \dfrac{\partial B}{\partial z} - \dfrac{n_0}{c}\dfrac{\partial B}{\partial t} = -iq^*(z)A - \dfrac{g}{2}B \end{cases} \tag{9.22}$$

9.2.1 Threshold conditions

We will derive the threshold conditions by performing a perturbative analysis of the coupled-mode equations 9.22. We will assume that the structure is not excited, i.e., there are no injected signals:

$$A(0,t) = 0 \qquad B(L,t) = 0 \tag{9.23}$$

If we now perturb the solution $A = B = 0$ of the coupled-mode equations, the threshold condition is reached when the perturbation grows with time. In the case of a real device, the perturbation is given by the spontaneous emission. We look for a solution as:

$$\begin{aligned} A(z,t) &= \widetilde{A}(z)e^{at} \\ B(z,t) &= \widetilde{B}(z)e^{at} \end{aligned} \tag{9.24}$$

Upon using Eqs. 9.24 in Eqs. 9.22 we obtain:

$$\begin{cases} \dfrac{d\widetilde{A}}{dz} = \left(\dfrac{g}{2} - \dfrac{n_0}{c}\,a\right)\widetilde{A} + iq(z)\widetilde{B} \\ \dfrac{d\widetilde{B}}{dz} = \left(-\dfrac{g}{2} + \dfrac{n_0}{c}\,a\right)\widetilde{B} - iq^*(z)\widetilde{A} \end{cases} \tag{9.25}$$

The general solution of the previous set of linear equations can be written as:

$$\begin{aligned} \widetilde{A}(z) &= M_{11}(z)\widetilde{A}(0) + M_{12}(z)\widetilde{B}(0) \\ B(z) &= M_{21}(z)\widetilde{A}(0) + M_{22}(z)\widetilde{B}(0) \end{aligned} \tag{9.26}$$

as will be shown in the next section in the case of a uniform grating and in the case of a $\lambda/4$-shifted grating. $M_{ij}(z)$ can be calculated when the coupling parameter $q(z)$ is known. Upon using the boundary conditions 9.23 in Eqs. 9.26 the following equations can be written:

$$\begin{cases} \widetilde{A}(L) = M_{12}(L)\widetilde{B}(0) \\ 0 = M_{22}(L)\widetilde{B}(0) \end{cases} \tag{9.27}$$

From the second equation we obtain:

$$M_{22}(L) = 0 \tag{9.28}$$

where, in general, $M_{22}(L)$ depends on the gain coefficient, g, and on the parameter a introduced in Eq. 9.24. Therefore, if the gain is fixed, a can be calculated: $a = a(g)$. It is clear that the applied perturbation increases with time only if $\mathrm{Re}(a) \geq 0$, indeed

$$A(z,t) = \widetilde{A}(z)e^{\mathrm{Re}(a)t}e^{i\mathrm{Im}(a)t}. \tag{9.29}$$

Threshold is reached when $\mathrm{Re}(a) = 0$. The corresponding value of g, which we will indicate as g_{th}, is the threshold gain. At threshold the parameter a can be written as:

$$a = -i\Omega = -i(\omega - \omega_0). \tag{9.30}$$

By using this result, the set of equations 9.25 can be re-written as:

$$\begin{cases} \dfrac{d\widetilde{A}}{dz} = i\Delta\kappa\widetilde{A} + iq(z)\widetilde{B} \\ \dfrac{d\widetilde{B}}{dz} = -i\Delta\kappa\widetilde{B} - iq^*(z)\widetilde{A} \end{cases} \tag{9.31}$$

where

$$\Delta\kappa = \frac{n_0}{c}\Omega - i\frac{g}{2} = \frac{n_0}{c}(\omega - \omega_0) - i\frac{g}{2} = \Delta\beta - i\frac{g}{2} = -i\left(\frac{g}{2} + i\Delta\beta\right). \tag{9.32}$$

In order to take into account the confinement factor in the active region, Eq. 9.32 can be re-written as follows:

$$\Delta\kappa = \frac{n_0}{c}(\omega - \omega_0) - i\Gamma\frac{g}{2}. \tag{9.33}$$

The general method illustrated in this section will be applied in the next section to derive the threshold gain and the frequencies of the oscillating modes of a uniform grating DFB laser and of a $\lambda/4$-shifted DFB laser.

Before proceeding, it is interesting to find the link between the perturbative analysis presented in this section and the discussion presented in Sect. 9.1, where the threshold conditions for laser action have been discussed in terms of the transmittance of the DFB device. Let us use the DFB structure as an amplifier: a signal is injected at $z = 0$ (at the frequency ω), while no signals are injected at $z = L$. From the set of Eqs. 9.26 we have:

$$\begin{cases} M_{11}(L)\widetilde{A}(0) + M_{12}(L)\widetilde{B}(0) = \widetilde{A}(L) \\ M_{21}(L)\widetilde{A}(0) + M_{22}(L)\widetilde{B}(0) = 0 \end{cases} \tag{9.34}$$

If we now divide both equations by $\widetilde{A}(0)$ we obtain:

$$\begin{cases} M_{11}(L) + rM_{12}(L) = t \\ M_{21}(L) + rM_{22}(L) = 0 \end{cases} \tag{9.35}$$

where we have introduced the field reflection, r, and transmission, t, coefficients:

$$\begin{aligned} r &= \left.\frac{\widetilde{B}(0)}{\widetilde{A}(0)}\right|_{\widetilde{B}(L)=0} \\ t &= \left.\frac{\widetilde{A}(L)}{\widetilde{A}(0)}\right|_{\widetilde{B}(L)=0} \end{aligned} \tag{9.36}$$

which can be easily calculated from Eqs. 9.35:

$$\begin{aligned} r &= -\frac{M_{21}(L)}{M_{22}(L)} \\ t &= \frac{\det(M)}{M_{22}(L)}, \end{aligned} \tag{9.37}$$

where $\det(M) = M_{11}M_{22} - M_{12}M_{21}$ is the determinant of the matrix $\{M_{ij}\}$. Since the structure is reciprocal (i.e., the transmission through the grating is equivalent for light incident from either side of the structure) $\det(M) = 1$, so that $t = 1/M_{22}$ and the grating transmittance can be written as

$$T = \frac{1}{|M_{22}(L)|^2}. \tag{9.38}$$

At threshold, with $\omega - \omega_0 = \Omega$, $M_{22}(L) = 0$ and the transmittance increases to infinity as we have seen, on a qualitative ground, in Sect. 9.1.

9.3 DFB laser with uniform grating

In the case of a uniform grating $\phi(z) = 0$ and $h(z) = 1$. The coupling parameter can be written as:

$$q(z) = q_0 = \frac{\delta n}{2n_0}\beta_0 = \frac{\delta n}{2n_0}\frac{\pi}{\Lambda} \tag{9.39}$$

The coupled-mode equations 9.31 can be solved by writing:

$$\widetilde{A}(z) = Ae^{\gamma z}, \qquad \widetilde{B}(z) = Be^{\gamma z} \tag{9.40}$$

The set of Eqs. 9.31 can be written as:

$$\begin{cases} (\gamma - i\Delta\kappa)A - iq_0 B = 0 \\ iq_0^* A + (\gamma + i\Delta\kappa)B = 0 \end{cases} \tag{9.41}$$

which can be easily solved. In order to have solutions different from the trivial one, $A = B = 0$, we must have:

$$\begin{pmatrix} \gamma - i\Delta\kappa & -iq_0 \\ iq_0^* & \gamma + i\Delta\kappa \end{pmatrix} = \gamma^2 + \Delta\kappa^2 - |q_0|^2 = 0 \tag{9.42}$$

which gives two eigenvalues:

$$\gamma_1 = \gamma = \sqrt{|q_0|^2 - \Delta\kappa^2}, \qquad \gamma_2 = -\gamma = -\sqrt{|q_0|^2 - \Delta\kappa^2} \tag{9.43}$$

The corresponding eigenvectors are:

$$\begin{pmatrix} iq_0 \\ \gamma - i\Delta\kappa \end{pmatrix}, \qquad \begin{pmatrix} iq_0 \\ -\gamma - i\Delta\kappa \end{pmatrix} \tag{9.44}$$

and the general solution can be written as:

$$\begin{pmatrix} \widetilde{A}(z) \\ \widetilde{B}(z) \end{pmatrix} = C_1 \begin{pmatrix} iq_0 \\ \gamma - i\Delta\kappa \end{pmatrix} e^{\gamma z} + C_2 \begin{pmatrix} iq_0 \\ -\gamma - i\Delta\kappa \end{pmatrix} e^{-\gamma z}. \tag{9.45}$$

When $z = 0$ we have:

$$\begin{cases} \widetilde{A}(0) = iq_0(C_1 + C_2) \\ \widetilde{B}(0) = \gamma(C_1 - C_2) - i\Delta\kappa(C_1 + C_2) \end{cases} \tag{9.46}$$

from which it is simple to calculate C_1 and C_2:

$$\begin{cases} C_1 = \dfrac{1}{2}\left(\dfrac{\widetilde{A}(0)}{iq_0} + \dfrac{\Delta\kappa}{\gamma}\dfrac{\widetilde{A}(0)}{q_0} + \dfrac{\widetilde{B}(0)}{\gamma}\right) \\ C_2 = \dfrac{1}{2}\left(\dfrac{\widetilde{A}(0)}{iq_0} - \dfrac{\Delta\kappa}{\gamma}\dfrac{\widetilde{A}(0)}{q_0} - \dfrac{\widetilde{B}(0)}{\gamma}\right) \end{cases} \tag{9.47}$$

So that we obtain:

$$\begin{cases} \widetilde{A}(z) = \left[\cosh(\gamma z) + i\dfrac{\Delta\kappa}{\gamma}\sinh(\gamma z)\right]\widetilde{A}(0) + \dfrac{iq_0}{\gamma}\sinh(\gamma z)\widetilde{B}(0) \\ \widetilde{B}(z) = -i\dfrac{q_0^*}{\gamma}\sinh(\gamma z)\widetilde{A}(0) + \left[\cosh(\gamma z) - i\dfrac{\Delta\kappa}{\gamma}\sinh(\gamma z)\right]\widetilde{B}(0) \end{cases} \tag{9.48}$$

which represent Eqs. 9.26 in the particular case of a uniform grating. The threshold condition is $M_{22}(L) = 0$:

$$\cosh(\gamma L) - i\frac{\Delta\kappa}{\gamma}\sinh(\gamma L) = 0 \tag{9.49}$$

which can be also written as:

$$\Delta\kappa\sinh(\gamma L) + i\gamma\cosh(\gamma L) = 0. \tag{9.50}$$

In the previous equation the real and imaginary components on the left hand side

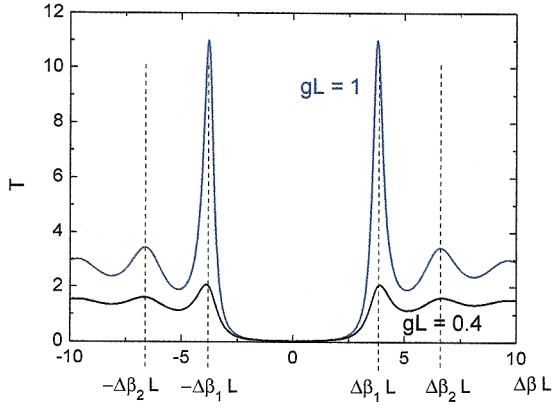

Figure 9.6: Intensity transmittance T of a uniform grating as a function of the detuning $\Delta\beta L$ calculated for two different values of the effective gain gL.

must be equal to zero, thus giving two equations. The unknown are the threshold gain, g_{th}, and the frequency, ω, of the oscillating mode. Therefore, the solution of Eq. 9.50 gives a set of (g_{th}, ω) values. From a practical point of view, to find the set of solutions (g_{th}, ω) corresponding to the various modes a particular gain value g is first fixed and the transmission $T = 1/|M_{22}(L)|^2$ is plotted as a function of the detuning $n_0\Omega/c$, as shown in Fig. 9.6. Below threshold the transmission presents various peaks, symmetrically placed with respect to $\Omega = 0$. Upon increasing the gain the first two maxima (at $\pm\Delta\beta_1$) increase and, at threshold, the transmission at these frequencies goes to infinity, thus giving the value of the threshold gain for the first two fundamental oscillating modes. In the same way the gain threshold for the modes at $\pm\Delta\beta_2$ can be determined by finding the gain value for which the two maxima at $\pm\Delta\beta_2$ increase to infinity.

We note that the threshold gain decreases upon increasing the value of the coupling parameter q_0. The stop-band of the structure is the frequency region between the two frequencies of the two lowest-order modes. In a real DFB laser with uniform grating only one of the two symmetric lowest-order modes can effectively oscillate, since structural imperfections, spatial hole burning and not equal anti-reflection coatings of the end faces can decrease the threshold gain of one of the two modes with respect to that of the other one. This is not the ideal situation in order to have a single mode oscillation, since it is difficult to determine which of the two lowest-order modes will eventually oscillate.

9.3.1 Approximated evaluation of oscillating frequencies and threshold gain

In this section we will calculate the oscillating frequencies and the threshold gain of a DFB with uniform grating in the case of strong gain, i.e., we will assume that $g \gg q_0$. Equation 9.50 can be re-written as follows:

$$e^{2\gamma L} = \frac{1+\gamma/(g/2+i\Delta\beta)}{1-\gamma/(g/2+i\Delta\beta)} \tag{9.51}$$

The term $\gamma/(g/2+i\Delta\beta)$ can be written in a different way, under the condition of strong gain:

$$\begin{aligned}\frac{\gamma}{g/2+i\Delta\beta} &= \frac{\sqrt{q_0^2+(g/2+i\Delta\beta)^2}}{g/2+i\Delta\beta} = \sqrt{1+\left(\frac{q_0}{g/2+i\Delta\beta}\right)^2} \\ &\approx 1+\frac{1}{2}\left(\frac{q_0}{g/2+i\Delta\beta}\right)^2\end{aligned} \tag{9.52}$$

By using this equation γ can be written as:

$$\gamma = \frac{g}{2} + i\Delta\beta + \frac{q_0^2(g/2-i\Delta\beta)}{2(g^2/4+\Delta\beta^2)}. \tag{9.53}$$

By using Eq. 9.52 in Eq. 9.51 and applying the strong-gain condition we have:

$$e^{2\gamma L} = -4\left(\frac{g/2+i\Delta\beta}{q_0}\right)^2. \tag{9.54}$$

which can be rewritten as follows:

$$\frac{1}{4}\left(\frac{q_0}{g/2+i\Delta\beta}\right)^2 e^{2\gamma L} = -1 \tag{9.55}$$

Upon taking the square root of both members of the previous equation we can write:

$$\frac{1}{2}\frac{q_0}{g/2+i\Delta\beta}e^{\gamma L} = \pm i \tag{9.56}$$

By equating the phases in Eq. 9.56 we obtain:

$$2\text{arctan}\left(\frac{2\Delta\beta}{g}\right) - 2\Delta\beta L + \frac{q_0^2\Delta\beta L}{(g/2)^2+\Delta\beta^2} = (2m+1)\pi \tag{9.57}$$

where $m = 0, \pm 1, \pm 2, \ldots$. In the limit $g \gg \Delta\beta, q_0$, the oscillating mode frequencies can be easily obtained from Eq. 9.57:

$$L\Delta\beta \approx -\left(m+\frac{1}{2}\right)\pi \tag{9.58}$$

The two lowest-order modes correspond to $m = 0$ and $m = -1$. When $m = 0$ the detuning is given by:

$$\Delta\beta = -\frac{\pi}{2L} \tag{9.59}$$

so that the oscillating frequency is:

$$\omega_{\omega=0} = \omega_0 - \frac{\pi c}{2n_0 L}. \tag{9.60}$$

When $m = -1$ the detuning is:

$$\Delta\beta = \frac{\pi}{2L} \tag{9.61}$$

and the corresponding oscillating frequency is:

$$\omega_{\omega=-1} = \omega_0 + \frac{\pi c}{2n_0 L}. \tag{9.62}$$

In the previous equations n_0 can be replaced by n_{eff}, the effective waveguide refractive index, so that:

$$\omega_m = \omega_0 \pm \frac{\pi c}{2n_{eff}L} \tag{9.63}$$

and the mode frequency spacing is given by:

$$\omega_m - \omega_{m-1} = \frac{\pi c}{n_{eff}L} \tag{9.64}$$

Therefore the two lowest-order modes are symmetyrically located with respect to the Bragg frequency ω_0.
By equating the amplitudes in Eq. 9.55 one can write:

$$\frac{e^{gL}}{(g/2)^2 + \Delta\beta^2} = \frac{4}{|q_0|^2}, \qquad (g = g_{th}) \tag{9.65}$$

which, by using Eq. 9.58 can be re-written as

$$e^{gL} = \frac{4}{(q_0 L)^2}\left[\left(\frac{gL}{2}\right)^2 + \left(m + \frac{1}{2}\right)^2 \pi^2\right], \qquad (g = g_{th}) \tag{9.66}$$

This equation can be solved graphically to obtain the threshold gain, as shown in

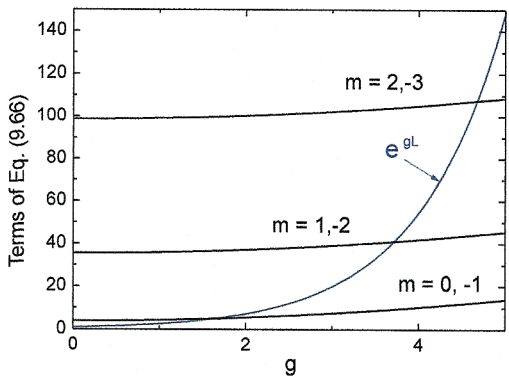

Figure 9.7: Graphical solution of Eq. 9.66. The intersections between the blue curve (which represents $\exp(gL)$) and the black curves (which represent the function on the right hand side of Eq. 9.66) give the threshold gain for the modes m. Note that the lowest-order modes ($m = 0$ and $m = -1$) have the same threshold gain.

Fig. 9.7. From Eq. 9.65 we see that, when q_0 is fixed, the threshold gain depends

only on $|\Delta\beta|$, therefore the two lowest-order modes ($m = 0, -1$) have the same threshold. Moreover, upon increasing m (therefore, upon increasing the detuning $\Delta\beta$), the threshold gain increases; indeed from Eq. 9.65 we have:

$$|q_0|^2 \frac{e^{gL}}{4} - \left(\frac{g}{2}\right)^2 = \Delta\beta^2, \tag{9.67}$$

so that:

$$\frac{|q_0|^2}{4} e^{gL} \approx \Delta\beta^2, \tag{9.68}$$

this is the reason for the spectral selectivity of DFB lasers.

9.4 DFB laser with $\lambda/4$-shifted grating

In this section we will consider DFB laser with a phase shift ϕ_0 at the center of the grating. In the case of a $\lambda/4-$shifted structure $\phi_0 = \pi$. We will consider a grating with a total length of $2L$, composed by two uniform gratings of length L, with a phase shift ϕ_0 between them. The first grating is characterized by a coupling parameter q_0 (given by Eq. 9.39), while the second grating is characterized by a coupling parameter $q_0 e^{i\phi_0}$. We have seen that in the case of a single uniform grating $\widetilde{A}(L)$ and $\widetilde{B}(L)$ can be calculated as follows:

$$\begin{pmatrix} \widetilde{A}(L) \\ \\ \widetilde{B}(L) \end{pmatrix} = \begin{pmatrix} \cosh(\gamma L) + i\frac{\Delta\kappa}{\gamma}\sinh(\gamma L) & i\frac{q_0}{\gamma}\sinh(\gamma L) \\ \\ -i\frac{q_0^*}{\gamma}\sinh(\gamma L) & \cosh(\gamma L) - i\frac{\Delta\kappa}{\gamma}\sinh(\gamma L) \end{pmatrix} \begin{pmatrix} \widetilde{A}(0) \\ \\ \widetilde{B}(0) \end{pmatrix}$$

In the case of two consecutive gratings the total matrix $\{M_{ij}\}$ is given by the product of the matrices of the two gratings:

$$\begin{aligned} M = & \begin{pmatrix} \cosh(\gamma L) + i\frac{\Delta\kappa}{\gamma}\sinh(\gamma L) & i\frac{q_0}{\gamma}\, e^{i\phi_0}\sinh(\gamma L) \\ \\ -i\frac{q_0^*}{\gamma}\, e^{-i\phi_0}\sinh(\gamma L) & \cosh(\gamma L) - i\frac{\Delta\kappa}{\gamma}\sinh(\gamma L) \end{pmatrix} \cdot \\ & \cdot \begin{pmatrix} \cosh(\gamma L) + i\frac{\Delta\kappa}{\gamma}\sinh(\gamma L) & i\frac{q_0}{\gamma}\sinh(\gamma L) \\ \\ -i\frac{q_0^*}{\gamma}\sinh(\gamma L) & \cosh(\gamma L) - i\frac{\Delta\kappa}{\gamma}\sinh(\gamma L) \end{pmatrix} \end{aligned} \tag{9.69}$$

The matrix element $M_{22}(2L)$ is given by:

$$M_{22}(2L) = \cosh^2(\gamma L) + \sinh^2(\gamma L)\left(\frac{|q_0|^2}{\gamma^2}\, e^{-i\phi_0} - \frac{\Delta\kappa^2}{\gamma^2}\right) - 2i\frac{\Delta\kappa}{\gamma}\cosh(\gamma L)\sinh(\gamma L) \tag{9.70}$$

The threshold condition can be obtained by equating $M_{22}(2L)$ to zero. In the case of a $\lambda/4$-shifted structure the threshold condition can be written as:

$$\left[\frac{q_0 \sinh(\gamma L)}{\gamma\cosh(\gamma L) - i\Delta\kappa\sinh(\gamma L)}\right]^2 = 1. \tag{9.71}$$

It is now possible to proceed as described in the case of uniform grating by plotting the transmittance $T = 1/|M_{22}^2(L)|^2$ of the structure for increasing values of the gain

coefficient. Laser threshold is reached when the transmittance increases to infinity. In this case, as already pointed out in Sect. 9.1, the lowest-order mode oscillates at the Bragg frequency: $\Delta\beta = 0$, $\beta = \beta_0 = \pi/\Lambda$. The higher-order modes are symmetrically located with respect to $\Delta\beta = 0$.

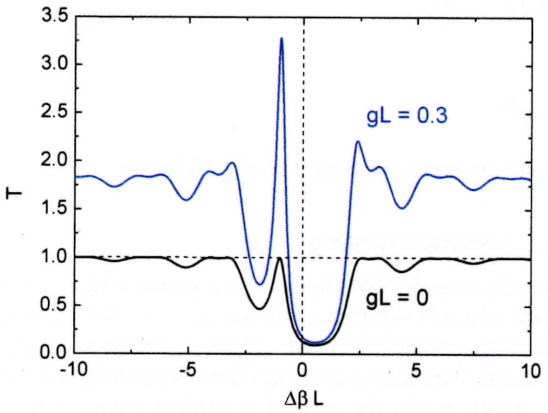

Figure 9.8: Intensity transmittance T of a shifted grating (with $\phi_0 = \pi/2$) as a function of the detuning $\Delta\beta L$ calculated for two different values of the effective gain gL.

The general expression for the matrix element $M_{22}(2L)$ given by Eq. 9.70 allows one to calculate the threshold gain and the mode frequencies when $\phi_0 \neq \pi$. As an example, Fig. 9.8 shows the transmittance of the structure calculated assuming $\phi_0 = \pi/2$. In this case the frequency of the lowest-order mode is $\beta < \beta_0$.

9.5 Distributed Bragg Reflector (DBR) laser

In the case of a DFB laser the grating structure, which provides the coupling between forward and backward waves, is produced along the cavity length of the laser. In the case of a Distributed Bragg reflector (DBR) laser two Bragg reflectors are placed at the ends of the laser cavity. The Bragg reflectors are formed by two passive (i.e., not pumped) gratings, where the refractive index is periodically modulated. A schematic illustration of a DBR laser is shown in Fig. 9.9(a). An important advantage is that in this case the grating is achieved in a portion of the device which is physically separated from the active region: this characteristic simplify the production procedures. Nevertheless, the frequency selectivity of a DBR laser is typically lower than that of a DFB laser, due to the presence of various longitudinal modes of the Fabry-Perot cavity. As a matter of fact, due to the very short cavity length, typically a single mode is contained within the high-reflectivity band of the structure, but temperature variations may induce mode hopping: for this reason DBR lasers are less frequently used then DFB lasers.

We can investigate the performances of a DBR laser by using the coupled-mode equation approach used in the case of DFB lasers. A DBR consists of a uniform grating of length L with:

$$\Delta\kappa = \Delta\beta + i\frac{\alpha}{2}, \qquad \gamma = \sqrt{|q_0|^2 - \Delta\beta^2} \tag{9.72}$$

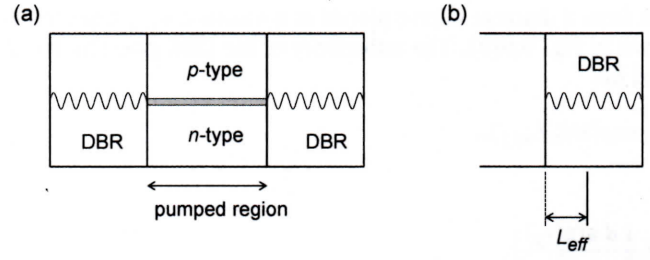

Figure 9.9: (a) Schematic structure of a DBR laser. (b) A DBR around the Bragg frequency can be modelled by a discrete mirror placed at a distance L_{eff} from the input of the DBR.

where α is the absorption coefficient. Its reflectivity can be calculated by using Eq. 9.48:

$$r_g = \frac{\widetilde{B}(0)}{\widetilde{A}(0)}\Big|_{\widetilde{B}(L)=0} = -\frac{M_{21}(L)}{M_{22}(L)} = \frac{iq_0^* \sinh(\gamma L)}{\gamma \cosh(\gamma L) - i\Delta\kappa \sinh(\gamma L)}, \tag{9.73}$$

The intensity reflectance of the DBR is given by $R = |r_g|^2$ and it is plotted in Fig.

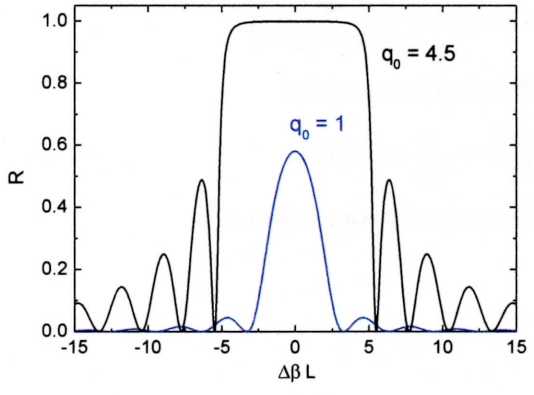

Figure 9.10: Intensity reflectance R of a DBR as a function of the detuning $\Delta\beta L$ calculated for two different values of the coupling parameter q_0.

9.10 as a function of the detuning $\Delta\beta L$ for two values of the coupling parameter q_0. The reflectivity is maximum when $\Delta\beta = 0$, i.e., when the propagation constant β of the input wave is equal to the Bragg propagation constant $\beta_0 = \pi/\Lambda$ of the grating. The reflectivity at the Bragg frequency can be calculated from Eq. 9.73 (we assume that the absorption in the grating can be neglected, so that $\Delta\kappa = \Delta\beta = 0$ and $\gamma = q_0$):

$$r_g = i \tanh(q_0 L) \tag{9.74}$$

The intensity reflectance $R = \tanh^2(q_0 L)$ increases with $|q_0|L$. It is also simple to calculate an approximated expression for the effective length L_{eff} of the Bragg reflector. The reflection of a DBR around the Bragg frequency can be modelled by

the reflection from a discrete mirror placed at a distance L_{eff} from the input of the DBR, as shown in Fig. 9.9(b). The reflectivity of the DBR given by Eq. 9.73 can be approximated by:

$$r_g \approx r_0 \exp(2i\beta_0 L_{eff}), \tag{9.75}$$

where:

$$L_{eff} = \frac{1}{2}\frac{d\ \arg(r_g)}{d\beta}\Big|_{\beta_0} \tag{9.76}$$

The phase $\phi(\beta)$ of the DBR can be expanded with a Taylor series around β_0:

$$\phi(\beta) \approx \phi(\beta_0) + \frac{\partial\phi}{\partial\beta}\Big|_{\beta_0}(\beta - \beta_0) \tag{9.77}$$

From Eq. 9.73, assuming a real coupling parameter q_0, the phase of the DBR can be calculated as:

$$\phi_{DBR}(\beta) = \pi - \arctan\left[\frac{\gamma}{\Delta\beta\tanh(\gamma L)}\right]. \tag{9.78}$$

Around $\beta = \beta_0$, assuming $\alpha = 0$, $\gamma = q_0$ and Eq. 9.78 can be re-written as:

$$\phi_{DBR}(\beta) = \pi - \arctan\left[\frac{q_0}{\Delta\beta\tanh(q_0 L)}\right]. \tag{9.79}$$

The effective length is therefore given by:

$$L_{eff} = \frac{1}{2}\frac{\partial\phi_{DBR}}{\partial\beta} = \frac{1}{2q_0}\tanh(q_0 L), \tag{9.80}$$

where we have derived Eq. 9.79 with respect to β.

10. Vertical Cavity Surface-Emitting Lasers

10.1 Basic structure

In Ch. 8 we have evidenced that edge-emitting lasers typically oscillated on several modes and the output beam is characterized by an elliptical shape, with high divergence, unless special optical systems are employed. For a number of important applications (in particular when the laser beam must be coupled into a single-mode optical fibre) a different laser geometry can be employed: a vertical cavity is used, where the laser beam propagates in the direction perpendicular to the junction. These lasers are called *Vertical Cavity Surface-Emitting Lasers* (VCSELs). A remarkable char-

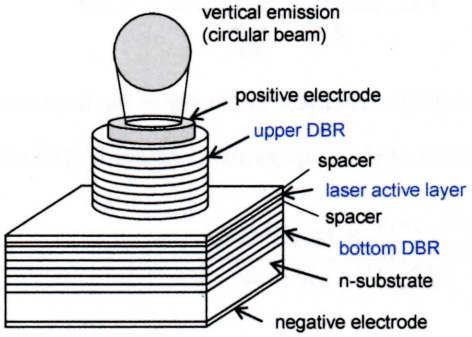

Figure 10.1: Schematic setup of a VCSEL.

acteristic of these lasers is the very short length of the active medium, of the order of the emission wavelength, which determines a very small gain. The laser cavity develops along the epitaxial growth axis of the semiconductor materials. The small gain requires end mirrors with very high reflectivities, in order to reduce internal

losses. In this case the Fresnel reflection at the semiconductor-air interface is by far not enough to guarantee sufficiently small losses in the cavity and Bragg reflectors are used, achieved by epitaxial deposition of alternating layers with low and high index of refraction with controlled thickness. A schematic structure of a VCSEL is shown in Fig. 10.1. Electrons and holes are injected into the device through the n-type (bottom DBR in Fig. 10.1) and p-type (upper DBR in Fig. 10.1) mirrors, respectively.

10.2 Threshold conditions

Let us first examine the threshold conditions for a VCEL. The threshold gain is given by Eq. 8.15, $\Gamma g_{th} = \gamma/L$, where L is the cavity length and Γ the confinement factor, which can be decomposed in a confinement factor in the transverse directions, x and y, and in an axial confinement factor in the direction z of beam propagation:

$$\Gamma = \Gamma_{xy}\Gamma_z. \tag{10.1}$$

$\gamma = \alpha L - \log R$ represents the cavity losses, where α is the absorption coefficient of the cavity and R is the average mirror reflectivity $R = \sqrt{R_1 R_2}$, where R_1 and R_2 are the reflectivities of the bottom and upper mirrors in Fig. 10.1.
The schematic geometry of the laser cavity is shown in Fig. 10.2. The standing-wave pattern of the electric field in the cavity can be written as:

$$E(x, y, z) = U(x, y)\sqrt{2}\cos\beta z, \tag{10.2}$$

where $U(x, y)$ is the normalized transverse profile of the electric field and $\beta = 2\pi\bar{n}/\lambda$

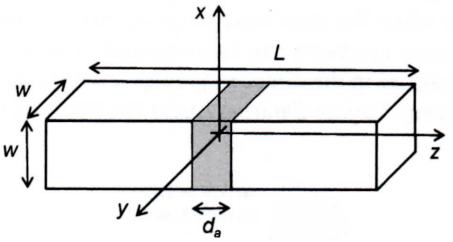

Figure 10.2: Schematic geometry of the laser cavity of a VCSEL.

is the z-component of the propagation constant $\mathbf{k}$. The confinement factors can be written as:

$$\Gamma_{xy} = \frac{\int_{-w/2}^{w/2}\int_{-w/2}^{w/2}|U(x,y)|^2 dxdy}{\int\int_{-\infty}^{\infty}|U(x,y)|^2 dxdy}, \tag{10.3}$$

$$\Gamma_z = \frac{1}{L}\int_{-d_a/2}^{d_a/2} 2\cos^2(\beta z)dz = \frac{d_a}{L}\left(1 + \frac{\sin(\beta d_a)}{\beta d_a}\right). \tag{10.4}$$

It is also useful to write the axial confinement factor Γ_z as the product of a fill factor $\Gamma_f = d_a/L$ and of an enhancement factor Γ_{enh} given by:

$$\Gamma_{enh} = 1 + \frac{\sin(\beta d_a)}{\beta d_a}. \tag{10.5}$$

Note that for thin active layers composed by a quantum well $\Gamma_{enh} \simeq 2$ and that it is typically in the range between 1 and 2. Here we have assumed that the active region is in the center ($z = 0$), in correspondence of the central maximum of the standing-wave pattern. If the active region is located at the field node, where $E(z) = 0$, the overlapping between the optical field and the active material is negligible and $\Gamma_{enh} \simeq 0$. Indeed, a more general expression of Γ_{enh} is the following:

$$\Gamma_{enh} = 1 + \cos(2\beta z_s)\frac{\sin(\beta d_a)}{\beta d_a}, \tag{10.6}$$

where z_s is the distance along the propagation direction between the center of the

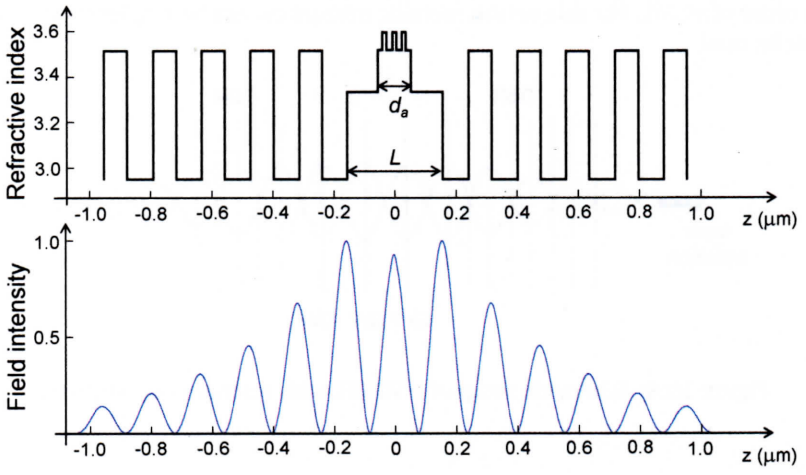

Figure 10.3: Upper panel: variation of the refractive index in the VCSEL described in the text. Lower panel: field intensity in the VCSEL.

active region and the peak of the standing-wave pattern inside the cavity. To illustrate this important concept, Fig. 10.3 shows an example of a VCSEL where the active material in composed by three InGaAs quantum wells, with a thickness of 8 nm, separated by 2 GaAs barriers (8-nm thick), cladded on both sides by 8-nm-thick GaAs layers (emission wavelength: 980 nm). The spacers are composed by $Al_{0.2}Ga_{0.8}As$. The DBRs are formed by alternating quarter-wavelength layers of AlAs and GaAs. The figure also displays the standing-wave pattern, $|E|^2$, inside the cavity. We will use this particular structure as a case study in this section. In this example the active region is placed in the center of the cavity in correspondence of the central anti-node of the standing-wave pattern ($z_s = 0$) so that the enhancement factor is of the order of 2. Moreover, the DBRs start with an AlAs quarter-wave layer close to the cavity. In this case at both ends of the inner cavity of length L the electric field presents a maximum amplitude and the resonance condition for the emission wavelength is given by:

$$\langle \bar{n} \rangle L = m\frac{\lambda_0}{2}, \tag{10.7}$$

where $\langle \bar{n} \rangle$ is the spatially averaged refractive index and m is a positive integer. Since the active layer must be located in an anti-node of the standing-wave pattern, the

minimum length of the cavity can be found from Eq. 10.7 with $m = 2$: $L = \lambda_0/\langle\bar{n}\rangle$. Note that the maximum total thickness of the active layer is of the order of $\lambda_0/4\langle\bar{n}\rangle$, since a larger thickness would correspond to active layers placed in regions with low optical intensity: this limits the number of quantum wells in the active region. In order to increase the gain it is also possible to use a particular scheme, called *resonant periodic gain*, which makes use of a relatively longer cavity, as shown in Fig. 10.4, where $m = 4$ so that the standing-wave pattern presents three anti-nodes and three quantum wells (or three double or triple quantum wells) separated by $\lambda_0/2\langle\bar{n}\rangle$ are placed at each anti-node.
Due to the very short length of the active medium, to achieve a reasonable threshold current the reflectivity of the bottom and upper mirrors must be very high, at least of the order of 99.5%. For this reason metallic mirrors cannot be employed and DBRs must be used.

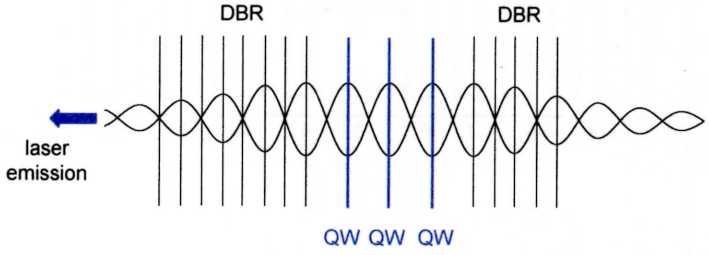

Figure 10.4: Schematic setup of a VCSEL with resonant periodic gain.

10.3 Distributed Bragg reflectors for VCSELs

The DBRs used in VCSELs are obtained by alternating quarter-wavelength thick layers with low and high refractive index, as schematically shown in Fig. 10.5. The essential idea is that at the Bragg wavelength the reflections at each index discontinuity add up in phase, thus leading to a large reflectivity. Generally more than 20 Bragg pairs are required for a mirror. In Fig. 10.5 r represents the reflectivity at each discontinuity and r_g represents the net mirror reflectivity, as in Sect. 9.5. If the wave is initially in the medium with refractive index n_2 and it is incident on the interface between medium 2 and 1 (at normal incidence), the electric field reflectivity is:

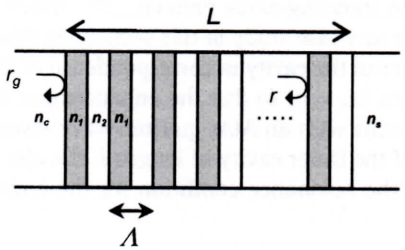

Figure 10.5: Schematic geometry of Bragg reflector.

$$r = \frac{n_2 - n_1}{n_2 + n_1}. \tag{10.8}$$

Note that if $n_2 < n_1$ the reflected field has a π phase shift with respect to the incident field. As already mentioned, the condition for constructive interference is obtained by alternating layers with thickness $d_1 = \lambda_0/4n_1$ and $d_2 = \lambda_0/4n_2$, where λ_0 is the free-space wavelength. At the Bragg wavelength the net electric field reflectivity of the mirror, assuming negligible absorption, can be written as:

$$R = \left(\frac{1-b}{1+b}\right)^2, \tag{10.9}$$

where b has two possible expressions. When the sequence of refractive indexes of the DBR shown in Fig. 10.5 is n_c (refractive index of the spacer layer), (n_1, n_2) repeated m times and finally n_s (refractive index of the substrate):

$$b = \frac{n_s}{n_c}\left(\frac{n_1}{n_2}\right)^{2m}, \tag{10.10}$$

while, when the refractive index sequence is n_c, (n_1, n_2) repeated m times, n_1 and finally n_s:

$$b = \frac{n_1^2}{n_c n_s}\left(\frac{n_1}{n_2}\right)^{2m}, \tag{10.11}$$

in this last case since the first and the last layers between the spacer and the substrate are made by the same material, the number of layer pairs of the DBR is $m + 0.5$. For example, if we consider a DBR formed by alternating layers of AlAs/GaAs and we assume that the waves are incident from $Al_{0.2}Ga_{0.8}As$ and are transmitted to air in the case of the top mirror and to GaAs in the case of the bottom mirror, the mirror reflectivities calculated according to the previous equations exceed 99.9% (neglecting absorption) for > 20 layer pairs.

The reflectivity of a Bragg mirror can be calculated as we have seen in Sect. 9.5. As shown in Fig. 9.10, when the number of layer pairs m is sufficiently high (corresponding to a high coupling parameter q_0) the mirror reflectivity is characterized by a broad spectral region of very high reflectivity around the Bragg wavelength λ_B. This region is called *stop-band* since it represents the wavelength interval where the periodic photonic structure does not allow transmission. The width of the stop-band is roughly given by[1]:

$$\Delta\lambda_{stop} \simeq \frac{2\lambda_B \Delta n}{\pi \langle n_g \rangle}, \tag{10.12}$$

where $\Delta n = |n_1 - n_2|$ and $\langle n_g \rangle$ is the spatial average of the group index given by Eq. 8.49. We can assume

$$\langle n_g \rangle = 2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1}. \tag{10.13}$$

For example in the case of a GaAs/AlAs mirror with $\Delta n = 0.57$ and $\langle n_g \rangle \simeq 3.2$ at $\lambda_B = 980$ nm, we obtain $\Delta\lambda_{stop} \approx 110$ nm.

As we have already observed in Sect. 9.5, in the case of a wave incident from a high index material, the mirror reflectivity can be written as:

$$r_g \simeq \sqrt{R} e^{2i\beta_0 L_{eff}}, \tag{10.14}$$

[1] see P. Yeh, *Optical Waves in Layered Media*, J. Wiley & Sons, New York, 1988.

so that a Bragg mirror can be modelled as an effective mirror with a field reflectivity amplitude equal to $|r_g|$ placed at a distance L_{eff} from the first layer of the DBR, as shown in Fig. 9.9(b). L_{eff} is given by Eq. 9.80, where L is the length of the DBR (see Fig. 10.5).
It is possible to shown that for a rectangular-shaped grating (i.e., for a grating where the modulation of the refractive index consists of a periodic square-wave function of period Λ) the coupling parameter can be written as:

$$q_0 = \frac{2\Delta n}{\lambda}\frac{\sin(\xi q\pi)}{q} \tag{10.15}$$

where q is the grating order and ξ the grating duty-cycle as shown in Fig. 10.6(a). If $\xi = 1/2$ and $q = 1$ the coupling parameter is given by:

$$q_0 = \frac{2\Delta n}{\lambda}. \tag{10.16}$$

In the case of low coupling (i.e., in the case of a DBR with low reflectivity) $\tanh(q_0 L) \approx q_0 L$ and the effective equivalent mirror is located at the center of the Bragg reflector $L_{eff} = L/2$, while in the case of high coupling $\tanh(q_0 L) \approx 1$ so that $L_{eff} = 1/2q_0 = \lambda/4\Delta n$. It is interesting to note that the effective length gives also an approximated value of the penetration depth of the optical energy into the Bragg reflector, defined as the depth into the mirror at which the optical field intensity is equal to $1/e$ of its value at the input of the mirror, as shown in Fig. 10.6(b).

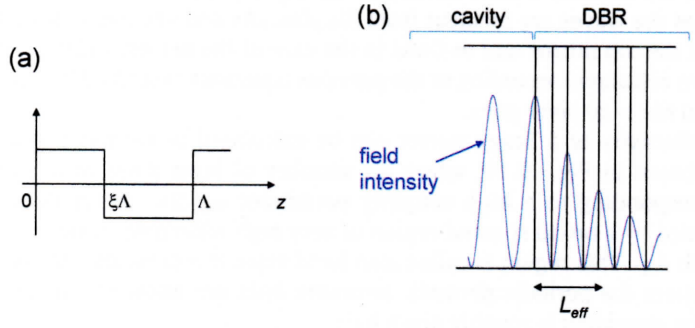

Figure 10.6: (a) Refractive index with a periodic square-wave modulation of period Λ and duty-cycle ξ. (b) Field intensity inside a Bragg mirror.

Taking into account the absorption losses inside a DBR, described by an absorption coefficient α_{DBR}, the mirror reflectivity can be written as:

$$R = R_m \exp(-2\alpha_{DBR} L_{eff}), \tag{10.17}$$

where R_m is the mirror reflectivity (with m pairs of layers) without absorption. Therefore, the maximum reflectivity (assuming $R_m = 1$) will be:

$$R_{max} = \exp\left(-\frac{\alpha_{DBR}\lambda_0}{2\Delta n}\right), \tag{10.18}$$

The total effective cavity length is given by:

$$L_{eff,T} = L + L_{eff,t} + L_{eff,b}, \tag{10.19}$$

where $L_{eff,t}$ and $L_{eff,b}$ are the effective penetration depth of the top and bottom DBRs, respectively. This quantity is used to estimate the spacing between consecutive longitudinal modes:

$$\Delta\lambda = \frac{\lambda^2}{2L_{eff,T}\langle n_g\rangle}. \tag{10.20}$$

Exercise 10.1 Calculate the spacing between longitudinal modes in the VCSEL shown in Fig. 10.3, with an emission wavelength of 980 nm. Assume that the cavity between the two Bragg mirrors is one optical wavelength measured in the semiconductor. The mirrors start with an AlAs quarter-wave layer adjacent to the cavity. Refractive indexes: for AlAs $n = 2.95$, for $Al_{0.2}Ga_{0.8}As$ $n = 3.39$, for GaAs $n = 3.52$, for InGaAs $n = 3.6$.

We can first calculate the thickness of the two $Al_{0.2}Ga_{0.8}As$ spacer layers. Between the two DBRs, the VCSEL is formed by two spacer layers, 4 GaAs layers and 3 InGaAs layers, so that we must have (the thicknesses are expressed in nanometers):

$$2 \cdot 3.39 \cdot s + 4 \cdot 3.52 \cdot 8 + 3 \cdot 3.6 \cdot 8 = 980 \text{ nm}$$

where $s = 115.2$ nm is the thickness of a spacer layer. The effective propagation length in both mirrors is (we assume a high coupling):

$$L_{eff} = \frac{\lambda_0}{4\Delta n} = 429.8 \text{ nm}$$

The thickness L of the cavity between the two mirrors is $L = 286.4$ nm. The total cavity effective length is therefore given by:

$$L_{eff.T} = (286.4 + 2 \cdot 429.8) \text{ nm} = 1146 \text{ nm}.$$

The longitudinal mode spacing is given by

$$\Delta\lambda = \frac{\lambda^2}{2L_{eff,T}\langle n_g\rangle} = 131 \text{ nm},$$

where $\langle n_g\rangle$ is the average index of refraction of the mirrors given by Eq. 10.13: $\langle n_g\rangle = 3.21$. The calculated mode spacing is larger than the stop-band of the mirror (it is also larger than the gain bandwidth of the quantum well used as active laser material). Therefore, just a single longitudinal mode can oscillate in this VCSEL. Depending on the lateral size of the VCSEL various transverse mode may oscillate and particular fabrication methods can be adopted to have a single transverse mode. ■

10.4 Threshold conditions and current confinement

The threshold gain can be written as follows:

$$\Gamma g_{th} = \langle\alpha_i\rangle + \frac{1}{L_{eff}} \log \frac{1}{\sqrt{R_1 R_2}}, \tag{10.21}$$

where R_1 and R_2 are the reflectivity of the top and bottom mirrors. Due to the small cavity length, absorption and scattering losses can be very small, thus leading to devices operating with low threshold current density (down to $\sim$ 350 A/cm^2). Since the transverse size is typically very small, in the case of single transverse mode operation (of the order of 5-10 μm) low threshold currents can be achieved. Typical threshold currents are of the order of 1 mA, with record values of about 0.06 mA.
In order to reduce the threshold current, it is crucial to use efficient current confinement structures. Figure 10.7 shows a few typical injection and current confinement schemes employed in VCSELs. The first one makes use of an annular electrode, which can limit the current flux near the ring electrode (see Fig. 10.7(a)). Laser radiation can be emitted from the inner part of the metallic ring. This structure is very simple but does not allow an efficient current confinement, mainly due to carrier diffusion. To limit carrier diffusion, proton implanted structures can be used (top or bottom emitting). In this case the current is confined by the semi-insulating region generated after the implant, which funnels into the active area (see Fig. 10.7(b)). Several commercial devices are based on this structure, which guarantees a good current confinement. Proton implantation may introduce optical losses and the implantation damage can affect the reliability of the laser. Optical losses introduced by this technique typically limit the laser efficiency to a maximum value of $\sim$ 20%. The ion implanted VCSEL allows a planar fabrication process. Both these schemes are examples of gain-guided VCSELs since the extension of the active region is determined by the current confinement, without any lateral photon confinement structures. In this case the carriers injected in the active region can diffuse, thus decreasing the overall laser efficiency. With low current injection, single transverse mode emission can be easily achieved.

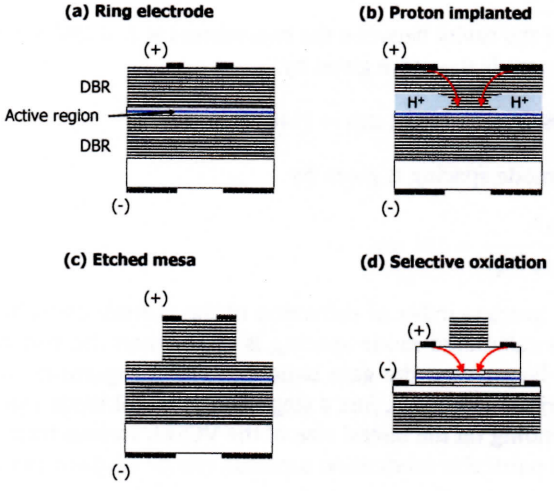

Figure 10.7: Typical current confinement schemes in VCSELs.

Another largely used configuration is the so-called *etched mesa* (see Fig. 10.7(c)), where current is confined to the mesa width. In this case the fabrication process in non-planar and the etch depth is of the order of a few micrometers. Mesa structures can be obtained by using various dry etching techniques (e.g., reactive ion etching, based on the use of $SiCl_4$, which can produce etch depths of the order of 10 μm;

chemically assisted ion-beam etching; etc.). The advantage of dry etching is that it allows the fabrication of structures with a very small diameter. Etching is stopped well above the active layer to maintain reliable device operation. The mesa structure provides an excellent current confinement and also an optical waveguide for the emitted photons. Another largely used structure is based on selective oxidation of a AlAs layer in the top Bragg reflector in order to generate a highly resistive layer of AlO_x just above the active layer (see Fig. 10.7(d)). This selective oxidation gives rise to an aperture in front of the active layer for the current flow and to a waveguide for the emitted photons. These devices are characterized by small optical losses thus leading to a very high efficiency ($> 50\%$) and very low threshold currents (<100 μA).

10.5 Advantages and applications

Since the first suggestion of a VCSEL device in 1977 and the demonstration of the first operation in 1979, VCSELs have become largely used laser devices for a wide range of applications, replacing the edge-emitting lasers in various applications where laser-optical fibre coupling is important, due to the small beam divergence and the symmetric mode profile. Coupling efficiency in single-mode optical fibres of the order of 90% can be achieved with top-emitting VCSELs. Emission wavelength and threshold conditions are relatively insensitive against temperature variations. Single longitudinal mode emission is easily achieved. A number of laser devices can be fabricated by fully monolithic processes thus leading to very low cost chip production. The VCSELs allow simple optical on-wafer testing, similar to the testing of integrated electronic circuits, directly after growth (i.e., before the wafer is cleaved). Moreover, densely packed two-dimensional laser arrays can be formed. Due to the short cavity round-trip time, VCSELs can be modulated with frequencies in the gigahertz range, with important applications for optical fibre communications.

11. Quantum cascade lasers

11.1 Quantum cascade lasers

In this final chapter we will analyse the quantum cascade lasers (QCLs), invented in 1994 by J. Faist, F. Capasso and co-workers[1], which are based on completely different physical processes compared to the semiconductor lasers we have considered so far. In the previous chapters we have always considered lasers where light emission in the active material takes place by means of recombination of an electron in the conduction band with a hole in the valence band (*interband transitions*). QCLs are unipolar devices, since they use only one type of charged carriers (the electrons) and the radiative recombination takes place within a single band (the conduction band) between two bound states of a potential well (*intraband transition*), as first theoretically suggested in 1971 by Kazarinov and Suris in a seminal paper[2]. Once an electron has undergone a quantum jump between two levels in a particular quantum well with the emission of a photon, it can tunnel into an adjacent well so that the emission process can be repeated: this process is called *electron recycling*. Therefore a single electron can produce the emission of multiple photons as it moves along the laser structure. This explains the name of this particular laser, since the emission is produced by a *cascade* process.

Figure 11.1 shows a simplified scheme of the conduction band structure for a basic QCL, where the laser transition is between sub-bands $|3\rangle$ and $|2\rangle$. The laser structure consists of consecutive stages composed by an electron injector and an active region. The number of stages typically ranges from 20 to 35 for lasers emitting in the wavelength region between 4 and 8 μm, but QCLs can have just a single stage or 100 stages. The tilt of the conduction band in Fig. 11.1 is produced by the applied electric field. The energy levels and wave-functions can be calculated

[1] J. Faist, F. Capasso, D.L. Sivco, C. Sirtori, A.L. Hutchinson, A.Y. Cho, "Quantum Cascade Laser," Science. **264**, 553 (1994).

[2] R.F. Kazarinov, R.A. Suris,"Possibility of amplification of electromagnetic waves in a semiconductor superlattice," Sov. Phys. Semicond. **5**, 707 (1971).

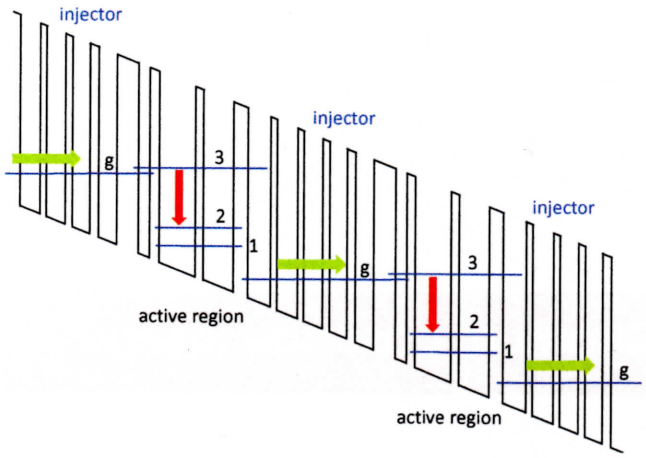

Figure 11.1: Schematic of the conduction band structure of a basic quantum cascade laser with laser transition between states $|3\rangle$ and $|2\rangle$ (red arrows). $|g\rangle$ is the ground state of the injector. The green arrows depict electron transport and injection into the active regions.

using the envelope function approximation. Electrons are injected into the upper laser level $|3\rangle$ by *resonant tunnelling*, which allows a highly selective injection, when the applied voltage exceeds a particular threshold value. In the particular structure shown in Fig. 11.1, which refers to an InGaAs/AlInAs laser, the thicknesses of the two quantum wells in the active region are 6.0 and 4.7 nm, separated by a 1.6-nm barrier, thus giving an energy separation of 207 meV (which corresponds to a wavelength of 6 μm) between the laser levels $|3\rangle$ and $|2\rangle$, and an energy separation between levels $|2\rangle$ and $|1\rangle$ of 37 meV, which is very close to the energy of an optical phonon mode of the material. In this way electrons in the lower laser level $|2\rangle$ rapidly scatter by emission of an optical phonon to the energy level $|1\rangle$, with a relaxation time $\tau_{21} = \tau_2 = 0.3$ ps. The lifetime of level $|3\rangle$ is significantly longer, due to the larger energy separation between levels $|3\rangle$ and $|2\rangle$, such that the electron-phonon scattering processes between these two states are non-resonant. Electrons injected into level $|3\rangle$ can scatter into levels $|2\rangle$ and $|1\rangle$ by phonon emission, with scattering times $\tau_{32} = 2.2$ ps and $\tau_{31} = 2.1$ ps, thus giving rise to an overall lifetime of the upper laser level $\tau_3 = (\tau_{31}^{-1} + \tau_{32}^{-1})^{-1} = 1.1$ ps, which is much longer than the lifetime of the lower laser level, thus allowing the creation of a population inversion between these two states. For this reason it is also important to prevent tunnelling from level $|3\rangle$ to states in the downstream adjacent injector region. This is obtained by proper engineering of the injector region so that at an energy corresponding to state $|3\rangle$ there is a region with very low density of states, called *minigap*, in the downstream injector. On the contrary, a dense manifold of states, called *miniband*,is present in the downstream injection region at energies corresponding to levels $|2\rangle$ and $|1\rangle$, in order to favour an efficient electron extraction from the active region after photon emission. The electrons injected in the downstream injection region gain in energy, relative to the bottom of the band and are injected into the following active region, so that the process described above proceeds in a stair-like cascade.

Due to the small energy difference between the laser levels $|3\rangle$ and $|2\rangle$, by proper engineering of the laser structure and materials it is possible to obtain laser emission in the mid- to far-infrared wavelength regions from 3.5 to 24 μm and even more, up to 60 μm. For wavelengths $\lambda > 5$ μm, wells and barriers are typically formed with alloy compositions lattice matched to the InP substrate (e.g., $Al_{0.48}In_{0.52}As$ and $In_{0.47}Ga_{0.53}As$). In this case the height of the quantum well barrier is the conduction band discontinuity between the two alloys, $\Delta\mathcal{E}_c = 0.52$ eV. For wavelengths shorter than 5 μm the energy of the upper laser state increases, thus reducing the barrier for thermal activation of electrons. This may be a limiting factor in cw operation at room temperature, since it may introduce temperature-induced instabilities. This problem can be alleviated by using strained AlInAs/GaInAs heterostructures. By taking advantage of the strain, the height of the quantum well barrier can be of the order of $0.7 - 0.8$ eV, thus suppressing the electron leakage over the barriers. Another limiting factor is the so-called *thermal back-filling* of the lower laser level by electrons from the downstream injector. In order to prevent a thermal activated electron transfer from the injector into level $|2\rangle$, the energy separation between this level and the injector ground state should be larger than 0.1 eV.

11.2 Gain coefficient

In Sect. 6.7 we have discussed the process of intersubband absorption; Eq. 6.87 gives the corresponding absorption/gain coefficient. In this equation the electronic densities N_i are measured in cm^{-3}; in the following we will use the carrier densities per unit area, $n_i = N_i L_p$, where L_p is the thickness of one period of active region and injector (such densities are therefore measured in cm^{-2}). In order to write the gain coefficient as a function of the photon energy, $\hbar\omega$, in Eq. 6.87, we have to write:

$$g(\omega - \omega_{mn}) = \hbar\, g(\hbar\omega - \mathcal{E}_m + \mathcal{E}_n) = \hbar\, g[\hbar(\omega - \omega_{nm})].$$

The squared modulus of the electric dipole moment, $|\mu_{nm}|^2$ can be written as $|\mu_{nm}|^2 = e^2|z_{nm}|^2$, where $z_{nm} = \langle n|z|m\rangle$ is the dipole matrix element between the states m and n. On the basis of Eq. 6.87 the gain coefficient can be written as:

$$g(\hbar\omega) = \frac{\omega\pi}{n_{eff}c\epsilon_0\hbar} e^2|z_{nm}|^2\hbar\, g_L[\hbar(\omega - \omega_{nm})]\frac{\Delta n}{L_p} \tag{11.1}$$

where n_{eff} is the mode refractive index, $\Delta n = n_m - n_n$ and $g_L(\hbar\omega)$ is the Lorentzian function:

$$g_L[\hbar(\omega - \omega_{nm})] = \frac{1}{\hbar}\frac{T_2/\pi}{1 + (\omega - \omega_{nm})^2 T_2^2} = \frac{\gamma_{nm}/\pi}{\hbar^2(\omega - \omega_{nm})^2 + \gamma_{nm}^2} \tag{11.2}$$

with a peak value given by $1/\pi\gamma_{nm}$ (where $\gamma_{nm} = \hbar/T_2$) and a full-width at half maximum given by $2\gamma_{nm}$. With these definitions, the peak gain/absorption can be written as:

$$g_p = \frac{2e^2\omega_{nm}}{n_{eff}c\epsilon_0 L_p}|z_{nm}|^2\frac{1}{2\gamma_{nm}}\Delta n = \sigma\Delta n. \tag{11.3}$$

To take into account the shape of the mode inside the cavity, the confinement factor Γ has to be used, so that the modal gain can be written as:

$$g_m = g_p\Gamma = \sigma\Gamma\Delta n. \tag{11.4}$$

We can also define a modal cross-section (measured in cm) as

$$\sigma_m = \frac{g_m}{\Delta n} = \Gamma\sigma, \tag{11.5}$$

which can be also written in terms of the emitted wavelength, λ:

$$\sigma_m = \frac{4\pi e^2}{n_{eff}\epsilon_0}\frac{|z_{nm}|^2}{2\gamma_{nm}L_p\lambda}\Gamma. \tag{11.6}$$

11.3 Rate equations and threshold conditions

In order to calculate the threshold gain and the threshold current density we will use a rate-equation approach as in Sect. 8.2, in the case of semiconductor lasers. We will model a QCL as a three-level system, as shown in Fig. 11.2, where the electron states $|3\rangle$ and $|2\rangle$ are the upper and lower laser levels, respectively, and n_i are the electron densities (in units of cm^{-2}). We will assume that electrons are injected into the upper laser level from the ground state $|g\rangle$ of the injector with an injection efficiency η; the remaining electrons scatter directly into the lower laser level.

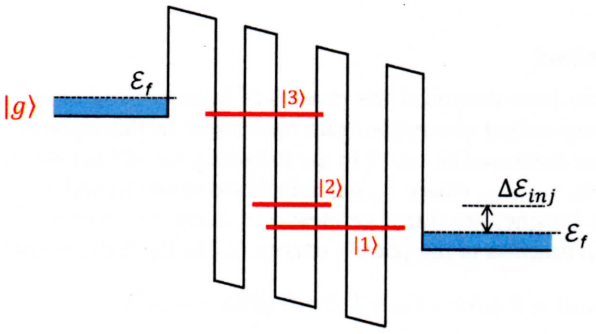

Figure 11.2: Three-level scheme of a QCL used in the rate equation approach.

$$\begin{cases} \dfrac{dn_3}{dt} = \eta\dfrac{J}{e} - \dfrac{n_3}{\tau_3} - R_{st} \\ \dfrac{dn_2}{dt} = (1-\eta)\dfrac{J}{e} + \dfrac{n_3}{\tau_{32}} + R_{st} - \dfrac{n_2 - n_2^{bf}}{\tau_2} \\ \dfrac{d\phi}{dt} = \Gamma R_{st} - \dfrac{\phi}{\tau_c} + \Gamma\beta_{sp}R_{sp} \end{cases} \tag{11.7}$$

where: J is the current density, $J = I/S$; ϕ is the photon density per unit area in the lasing mode (measured in cm^{-2}); R_{st} is the net stimulated emission rate. In the following we will neglect the spontaneous emission term, $\Gamma\beta_{sp}R_{sp}$. The stimulated emission term, R_{st}, is given by Eq. 8.5:

$$R_{st} = v_g g\phi = \frac{c}{n_{eff}}\sigma(n_3 - n_2)\phi \tag{11.8}$$

The last term in the second rate equation takes into account the process of thermal back-filling from the injector back into the lower laser level $|2\rangle$ mentioned in Sect. 11.1. If n_s is the sheet carrier density in the injector ground state, n_2^{bf} is given by

$$n_2^{bf} = n_s \exp\left(-\frac{\Delta\mathcal{E}_{inj}}{k_B T}\right), \tag{11.9}$$

where $\Delta\mathcal{E}_{inj}$ is the energy difference between level $|2\rangle$ and the quasi-Fermi level in the injector, as shown in Fig. 11.2.

We will assume steady-state conditions, so that the temporal derivatives in the rate equations are set to zero. At threshold $\phi = 0$ and, from the first rate equation, we have:

$$n_3 = \eta\tau_3\frac{J}{e}, \tag{11.10}$$

which is then used in the second rate equation, where $n_2 = n_3 - \Delta n$, to calculate the population inversion $\Delta n = n_3 - n_2$:

$$\Delta n = \frac{J}{e}[\eta\tau_{eff} - (1-\eta)\tau_2] - n_2^{bf}, \tag{11.11}$$

where

$$\tau_{eff} = \tau_3\left(1 - \frac{\tau_2}{\tau_{32}}\right) \tag{11.12}$$

is the effective lifetime of the population inversion. Indeed, assuming $\eta = 1$, from Eq. 11.11 we have

$$\frac{d\Delta n}{\tau_{eff}} = \frac{dJ}{e}. \tag{11.13}$$

At threshold the spatially averaged gain must equal the spatially averaged total losses, which we will indicate as α_{tot}. The population inversion is clamped at the threshold value, Δn_{th}. From Eq. 8.15 we can write:

$$\Gamma\sigma\Delta n_{th} = \sigma_m\Delta n_{th} = \alpha_{tot} \tag{11.14}$$

By using Eq. 11.11 in the previous equation, we can calculate the threshold current density, J_{th}:

$$J_{th} = \frac{e(\alpha_{tot}/\sigma_m + n_2^{bf})}{\eta\tau_{eff} - (1-\eta)\tau_2}. \tag{11.15}$$

If we assume a unit injection efficiency ($\eta = 1$), we obtain:

$$J_{th} = \frac{e}{\tau_{eff}}\left(\frac{\alpha_{tot}}{\sigma_m} + n_2^{bf}\right) = \frac{e}{\tau_{eff}}(\Delta n_{th} + n_2^{bf}). \tag{11.16}$$

From this formula it is clear that before reaching the threshold condition the current has to compensate both the losses inside the cavity and the electrons injected into level $|2\rangle$ by thermal back-filling. This last term strongly depends on temperature, as shown in Eq. 11.9. To minimize n_2^{bf} compared to α_{tot}/σ_m a sufficiently high energy difference $\Delta\mathcal{E}_{inj}$ (typically > 0.1 eV) should be achieved by proper bandgap engineering.

If we assume $\eta = 1$ and the back-filling process is neglected, from Eq. 11.11 the population inversion is given by:

$$\Delta n = \tau_{eff}\frac{J}{e}, \tag{11.17}$$

and the modal gain can be written as:

$$g_m = \sigma_m \Delta n = \left(\frac{4\pi e}{n_{eff}\epsilon_0} \frac{|z_{32}|^2}{2\gamma_{32} L_p \lambda} \right) \tau_3 \left(1 - \frac{\tau_2}{\tau_{32}} \right) \Gamma J_{th} \equiv g(\lambda) \Gamma J_{th}, \tag{11.18}$$

where we have introduced the gain coefficient $g(\lambda)$. With this definition, the threshold condition given by Eq. 11.14 can be written as:

$$g(\lambda)\Gamma J_{th} = \alpha_{tot}. \tag{11.19}$$

It is evident that a crucial parameter for the determination of the threshold conditions is the total cavity loss α_{tot}. A first contribution to α_{tot} is related to the mirror losses:

$$\alpha_m = \frac{1}{2L} \log \left(\frac{1}{R_1 R_2} \right), \tag{11.20}$$

where R_1 and R_2 are the reflectivity of the two laser end mirrors. A source of cladding losses is related to free-carrier absorption in the doped semiconductor regions: we will indicate this loss contribution with α_{fc}. Resonant absorption is taken into account by the thermal back-filling factor n_2^{bf}. Non-resonant losses (α_{nr}) of the active region are due to all the other transitions in the active region. Scattering losses (α_{sc}) are due to light-scattering processes in the cavity mainly related to waveguide sidewall roughness. Therefore we can write:

$$\alpha_{tot} = \alpha_m + \Gamma\alpha_{nr} + (1 - \Gamma)\alpha_{fc} + \alpha_{sc} = \alpha_m + \alpha_w. \tag{11.21}$$

The threshold current density given by Eq. 11.16 can be written as

$$J_{th} = \frac{1}{\tau_{eff}} \left[\frac{n_{eff}\epsilon_0 2\gamma_{32} L_p \lambda}{4\pi e |z_{32}|^2} \frac{\alpha_m + \alpha_w}{\Gamma} + e n_s \exp\left(-\frac{\Delta\mathcal{E}_{inj}}{k_B T} \right) \right]. \tag{11.22}$$

This formula can be improved to reproduce the temperature dependence of J_{th} by considering temperature dependent lifetimes, τ_i ($i = 1, 2, 3$):

$$\tau_i(T) = \tau_{i0} \frac{1}{1 + 2/\left[\exp\left(\mathcal{E}_{LO}/k_B T\right) - 1\right]}, \tag{11.23}$$

where τ_{i0} is the scattering time at low temperature and $\mathcal{E}_{LO}$ is the energy of the longitudinal optical (LO) phonons.

A simple formula giving the temperature dependence of the threshold current density can be written, as in conventional semiconductor lasers, by introducing a characteristic temperature T_0:

$$J_{th}(T) = J_0 \exp\left(\frac{T}{T_0} \right). \tag{11.24}$$

Typical values of T_0 are in the temperature range between 100 and 200 K. A slightly different version of Eq. 11.24, which allows one to fit the dependence of J_{th} versus temperature over the entire temperature range of laser operation (from 4.2 K to above the room temperature) is the following:

$$J_{th}(T) = J_0 \exp\left(\frac{T}{T_0} \right) + J_1. \tag{11.25}$$

11.4 Output power, slope- and wall-plug efficiency

Above threshold the population inversion is clamped to its threshold value, $\Delta n = \Delta n_{th}$, and the photon density, ϕ, is no longer zero. From the rate equations 11.7 it is possible to calculate ϕ (we will neglect spontaneous emission). For simplicity we will assume $\eta = 1$. From the first two rate equations we have:

$$R_{st} = \frac{\tau_{eff}}{\tau_2 + \tau_{eff}} \frac{J - J_{th}}{e} \tag{11.26}$$

where J_{th} is given by Eq. 11.16. From the third rate equation we have $\phi = \Gamma \tau_c R_{st}$, so that:

$$\phi = \frac{\tau_c \tau_{eff}}{\tau_2 + \tau_{eff}} \frac{\Gamma}{e} (J - J_{th}), \tag{11.27}$$

where the photon lifetime is given by:

$$\tau_c = \frac{n_{eff}}{c\alpha_{tot}} \tag{11.28}$$

so that the photon density (per unit area) can be written as:

$$\phi = \frac{\tau_{eff} \Gamma (J - J_{th}) n_{eff}}{e\alpha_{tot} c(\tau_2 + \tau_{eff})}. \tag{11.29}$$

To calculate the output power we will proceed as in the case of conventional semiconductor lasers. The optical energy stored in the laser cavity can be written as

$$\mathcal{E}_{st} = \hbar\omega (N_p S_p) \phi, \tag{11.30}$$

where N_p is the number of periods (i.e., the number of stages of active regions and injectors), $\hbar\omega$ the energy of the emitted photons and the surface S_p is related to the volume V_p occupied by photons (typically larger than the volume of the active region occupied by electrons, $V = \Gamma V_p$). The photon loss rate due to mirror transmission is given by:

$$\frac{c}{n_{eff}} \frac{1}{L} \log\left(\frac{1}{R}\right) = \alpha_m \frac{c}{n_{eff}} \tag{11.31}$$

so that the output power is given by:

$$P = \mathcal{E}_{st} \alpha_m \frac{c}{n_{eff}} = \frac{\hbar\omega}{e} N_p \frac{\alpha_m}{\alpha_{tot}} \frac{\tau_{eff}}{\tau_2 + \tau_{eff}} (I - I_{th}) \tag{11.32}$$

where we have used the current $I = J\Gamma S_p$. The slope efficiency is:

$$\eta_s = \frac{dP}{dI} = \frac{\hbar\omega}{e} N_p \frac{\alpha_m}{\alpha_{tot}} \frac{\tau_{eff}}{\tau_2 + \tau_{eff}} = \eta_{int} N_p \frac{\hbar\omega}{e} \frac{\alpha_m}{\alpha_{tot}} \tag{11.33}$$

where we have introduced the *internal quantum efficiency* η_{int}:

$$\eta_{int} = \frac{\tau_{eff}}{\tau_2 + \tau_{eff}}. \tag{11.34}$$

It is evident that upon increasing the number of periods, N_p, large slope efficiencies and therefore large output powers can be achieved: values in the range of W/A and W, respectively, are typical. The external quantum efficiency, η_{ext} is given by:

$$\eta_{ext} = \frac{d(P/\hbar\omega)}{d(I/e)} = \eta_s \frac{e}{\hbar\omega} = \eta_{int} N_p \frac{\alpha_m}{\alpha_{tot}} \tag{11.35}$$

so that the output power can be calculated as:

$$P = \eta_{ext} \frac{\hbar\omega}{e}(I - I_{th}) \tag{11.36}$$

as in the case of conventional semiconductor lasers.

11.4.1 Wall-plug efficiency

The wall-plug efficiency, which gives the conversion efficiency between electrical power, P_e, and optical power, P, is defined as:

$$\eta_{wp} = \frac{P}{P_e} = \frac{P}{UI}, \tag{11.37}$$

where U is the voltage drop on the structure. In the case of a QCL an equivalent and useful definition of η_{wp} is:[3]

$$\eta_{wp} = \frac{dP}{UdI}\frac{J_{max} - J_{th}}{J_{max}} = \frac{\eta_s}{U}\frac{J_{max} - J_{th}}{J_{max}}, \tag{11.38}$$

where J_{max} is the maximum current density. So far we have assumed that electrons are injected into the upper laser level $|3\rangle$ from the preceding injection region quickly and almost exclusively and that they are efficiently removed from states $|2\rangle$ and $|1\rangle$. Electrons are injected into level $|3\rangle$ by *resonant tunneling* and the electron injection into the active region is optimized when quasi-equilibrium is reached between the population of the injector ground state and that of the upper laser level, which are therefore characterized by a common quasi-Fermi level. In this way space-charge build-up in the active region induced by increased electron injection is avoided. A very good expression for the maximum current density is given by the following expression:

$$J_{max} = n_s \frac{e}{\tau_t} \tag{11.39}$$

where n_s is the free carrier sheet density per period (i.e., the 2D electron concentration in the injector state) and τ_t is an effective transit time of the electron across a period of the active region at resonance. The concept of resonance is illustrated in Fig. 11.3, which shows the process of injection through a very thin barrier from the ground state of the injector, $|g\rangle$, to the upper laser level at resonance (the corresponding injection into the lower laser level is much less efficient). The resonance condition is obtained by applying an electric field, E_{res}, to the structure, which decreases the energy detuning between $|g\rangle$ and $|3\rangle$ states until it vanishes and the states $|g\rangle$ and $|3\rangle$ form a doublet with anti-crossing energy $2\hbar|\Omega|$ (as shown in Fig. 11.3). The voltage

[3] J. Faist, "Wallplug efficiency of quantum cascade lasers: Critical parameters and fundamental limits," Appl. Phys. Lett. **90**, 253512 (2007).

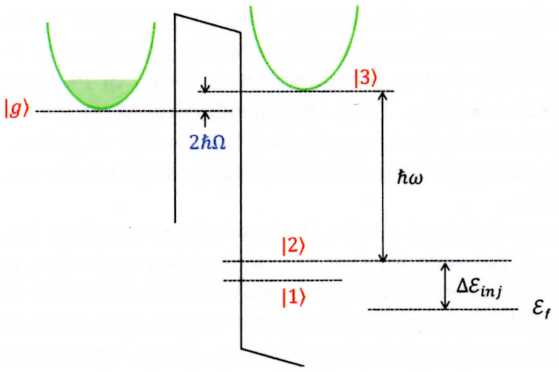

Figure 11.3: Schematic representation of the resonant tunneling process from the injector to the upper laser level.

drop on a single period is given, as depicted in Fig. 11.3, by $eU_p = \hbar\omega + \Delta\mathcal{E}_{inj}$, so that the voltage drop on the complete structure with N_p periods is given by:

$$U = \frac{N_p}{e}(\hbar\omega + \Delta\mathcal{E}_{inj}). \tag{11.40}$$

The threshold current density in the case of a N_p-period structure is obtained from Eq. 11.16 by using the total modal gain cross-section $N_p\sigma_m$:

$$J_{th} = \frac{e}{\tau_{eff}}\left(\frac{\alpha_{tot}}{N_p\sigma_m} + n_2^{bf}\right). \tag{11.41}$$

By using Eqs. 11.35, 11.39-11.41 in Eq. 11.38, the wall-plug efficiency can be written as:

$$\eta_{wp} = \eta_{int}\frac{\alpha_m}{\alpha_{tot}}\frac{1}{1+\Delta\mathcal{E}_{inj}/\hbar\omega}\left[1 - \frac{\tau_t}{\tau_{eff}}\left(\frac{\alpha_{tot}}{n_s N_p \sigma_m} + \frac{n_2^{bf}}{n_s}\right)\right]. \tag{11.42}$$

From Eq. 11.21, $\alpha_{tot} = \alpha_m + \alpha_w$. It has been shown that for QCLs operating both in the mid- and far-infrared spectral region, the waveguide loss, α_w, is proportional to the doping and can be written as:

$$\alpha_w = N_p\alpha_{fc}n_s + \alpha_{sc}. \tag{11.43}$$

In good quality lasers, α_{sc} can be neglected. Moreover we will assume that, by proper laser design, $\Delta\mathcal{E}_{inj}$ is large enough so that back-filling process can be neglected. By using these assumptions it is simple to calculate the value of the mirror losses, α_m, which maximises the wall-plug efficiency:

$$\alpha_m = \sqrt{N_p\sigma_m n_s \alpha_w \tau^*} - \alpha_w, \tag{11.44}$$

where we have introduced the dimensionless upper state lifetime $\tau^* = \tau_{eff}/\tau_t$. By using the previous expression in Eq. 11.42 (with $n_2^{bf} = 0$) it is possible to calculate the maximum value of the wall-plug efficiency:

$$\eta_{wp,max} = \eta_{tr}\frac{1}{1+\Delta\mathcal{E}_{inj}/\hbar\omega}\left[\frac{\sqrt{g^*\tau^*}-1}{\sqrt{g^*\tau^*}}\right]^2 \tag{11.45}$$

where we have introduced the dimensionless gain cross-section $g^* = \sigma_m/\alpha_{fc}$.

It is interesting to note that the first QCL demonstrated by Faist, Capasso and coworkers in 1994 was characterized by a wall-plug efficiency $< 0.15\%$ in pulsed operation at 10 K and that η_{wp} remained of the order of 1% at room temperature for most QCLs for a decade. Then, the introduction of optimized laser structures allowed a significant increase of η_{wp} up to $\sim 50\%$. A first optimization consists in the reduction of the voltage drop in the active region, which does not contribute to the generation of light.
Another technique is based on the increase of the coupling strength, $\hbar\Omega$, since the tunneling rate between injector and active layer and other parameters which determine the laser efficiency, strongly depends on this parameter. The increase of the tunneling rate gives rise to an increase of the maximum current density, J_{max}, thus leading to an increase of the wall-plug efficiency. A large tunneling rate presents other advantages: it reduces the electron density in the injector region, thus decreasing the leakage current from the injector to the lower laser level or the continuum energy levels; moreover, it increases the population inversion and therefore the gain.

The effects of the coupling strength on the electron transport characteristics can be understood by using the following expression (reported by Kazarinov and Suris) for the current between two sub-bands coupled by a tunneling barrier with a coupling energy $2\hbar|\Omega|$ (see Fig. 11.3):

$$J = en_s \frac{2|\Omega|^2 \tau_\perp}{1 + \Delta^2 \tau_\perp^2 + 4|\Omega|^2 \tau_3 \tau_\perp}, \tag{11.46}$$

where $\tau_\perp$ is the relaxation time of electron in-plane momentum, $\hbar\Delta$ is the energy level detuning from full resonance, which is related to the applied electric field E by the following relation:

$$\hbar\Delta = ed(E - E_r), \tag{11.47}$$

where d is the effective spatial distance between resonantly coupled injector state and upper laser level state, and E_r is the electric fiel required for full resonance. From Eq. 11.46 it is evident that upon increasing $|\Omega|$, the detuning term in the denominator, $\Delta^2\tau_\perp^2$, does not significantly influence the resonant tunneling rate.
It has been recognized that interface roughness introduces a significant detuning to the energy levels in resonance (interfaces between adjacent semiconductor layers present roughness of the order of a few ångstroms, which is significant since typical semiconductor layer thicknesses in QCLs are of the order of $10 - 50$ Å). An increase of the coupling strength can be achieved by using very thin injector barriers, of the order of $\sim$ 10 Å, between the injector and the active region. The obtained high coupling strength reduces the relative importance of the detuning of resonant tunneling induced by the interface roughness. Another advantage is that the upper laser level spreads even more into the injection region as a consequence of the very thin injector barrier: this effect increases the lifetime of the upper laser level and decreases the threshold current density. Detailed calculations based on a density matrix approach[4] show that the gain coefficient rapidly increases upon increasing the energy splitting $2\hbar|\Omega|$ up to $\sim$ 20 meV and then it slowly decreases.

[4]see J.B. Khurgin *et al.*, "Role of interface roughness in the transport and lasing characteristics of quantum-cascade lasers," Appl. Phys. Lett. **94**, 091101 (2009)

11.5 Numerical example

In this numerical example we will refer to a particular QCL described by Gmachl and co-workers in Rep. Prog. Phys. **64**, 1533 (2001). The active region is composed by 3 InGaAs quantum wells closely coupled by thin AlInAs barriers. The energy separation between the upper and lower laser levels is $\mathcal{E}_{32} = 153.6$ meV (which corresponds to an emission wavelength $\lambda = 8.08$ μm) at the electric field value $E = 45$ kV cm^{-1}, corresponding to the measured threshold. The energy separation between the levels $|2\rangle$ and $|1\rangle$ is $\mathcal{E}_{21} = 38.3$ meV, designed to have an efficient depletion of the lower laser level by resonant LO-phonon scattering (indeed $\mathcal{E}_{LO} \approx 34$ meV). The energy separation between the upper laser level and the ground state of the injector is $\mathcal{E}_{3g} = 40$ meV. The structure is designed to have level $|1\rangle$ of the preceding active region in resonance with level $|3\rangle$ of the following active region in order to have a resonant carrier transport between consecutive active regions without a significant carrier relaxation in the injector. LO-phonon scattering time from level $|3\rangle$ to level $|2\rangle$ is $\tau_{32} = 3.1$ ps, from level $|3\rangle$ to level $|1\rangle$ is $\tau_{31} = 3.6$ ps, from level $|3\rangle$ to all states of the downstream injection region is $\tau_{3i} = 14.6$ ps (i.e., $\tau_{3i} \gg \tau_{31}, \tau_{32}$). Lifetime of state $|2\rangle$ is $\tau_2 \approx \tau_{21} = 0.3$ ps. The dipole matrix element of the optical transition is $|z_{32}| = 1.9$ nm. The sheet carrier density in the injector is $n_s = 1.6 \times 10^{11}$ cm^{-2}. The full-width at half maximum $2\gamma_{32}$ (measured value of the luminescence linewidth) is $2\gamma_{32} \approx 10$ meV. The length of one period is $L_p = 44.3$ nm, the effective refractive index of the waveguide is $n_{eff} = 3.27$, the number of periods is $N_p = 30$, the facet reflectivity is $R = 0.28$ and the length of the laser is $L = 2.25$ mm. The measured waveguide losses are $\alpha_w = 24$ cm^{-1}. The confinement factor, Γ, for the complete structure with N_p periods is $\Gamma = 0.5$. We assume an energy separation between the ground state of the injector and state $|2\rangle$ of the preceding active region $\Delta\mathcal{E}_{inj} = 150$ meV. Coupling energy $\hbar\Omega = 1.3$ meV, in-plane momentum relaxation time $\tau_\perp = 40$ fs. Calculate the gain coefficient $g(\lambda)$, the threshold current density, J_{th}, the slope efficiency per facet, the external quantum efficiency per facet per stage and an approximated value of the wall-plug efficiency.

The lifetime of the upper laser level is given by:

$$\tau_3 = \left(\frac{1}{\tau_{31}} + \frac{1}{\tau_{32}} + \frac{1}{\tau_{3i}}\right)^{-1} = 1.5 \text{ ps}.$$

The gain coefficient, $g(\lambda)$, can be calculated by using Eq. 11.18:

$$g(\lambda) = \left(\frac{4\pi e}{n_{eff}\epsilon_0} \frac{|z_{32}|^2}{2\gamma_{32} L_p \lambda}\right) \tau_3 \left(1 - \frac{\tau_2}{\tau_{32}}\right) = 59.1 \text{ cm/kA}.$$

The mirror loss α_m is given by:

$$\alpha_m = -\frac{1}{L} \log R = 5.6 \text{ cm}^{-1},$$

the threshold current density is (from Eq. 11.19):

$$J_{th} = \frac{\alpha_m + \alpha_w}{\Gamma g(\lambda)} = 1 \text{ kA/cm}^2,$$

in excellent agreement with the experimental value $J_{th,exp} \approx 1$ kA/cm^2. The effective lifetime τ_{eff} is given by:

$$\tau_{eff} = \tau_3 \left(1 - \frac{\tau_2}{\tau_{32}}\right) = 1.35 \text{ ps}.$$

The internal quantum efficiency can be calculated by using Eq. 11.34:

$$\eta_{int} = \frac{\tau_{eff}}{\tau_2 + \tau_{eff}} = 0.82.$$

The slope efficiency per facet can be calculated from Eq. 11.35:

$$\eta_s = \eta_{int} N_p \frac{\hbar\omega}{e} \frac{\alpha_m/2}{\alpha_w + \alpha_m/2} = 394 \text{ mW/A},$$

again in excellent agreement with the experimental value $\eta_{s,exp} \approx 400$ mW/A. The external quantum efficiency per facet and per stage is given by:

$$\frac{\eta_{ext}}{N_p} = \frac{e}{\hbar\omega} \frac{\eta_s}{N_p} = 8.5\%.$$

Therefore, external quantum efficiencies larger than 1 for the complete laser structure can be achieved by using a number of periods $N_p \geq 12$.
In order to calculate an approximated value of the wall-plug efficiency we will first estimate the effective transit time of the electrons across a period of the active region at resonance, τ_t, by using Eq. 11.46 and Eq. 11.39:

$$\tau_t \approx \frac{1 + 4\Omega^2 \tau_3 \tau_\perp}{2\Omega^2 \tau_\perp},$$

where we have assumed zero detuning from the resonance condition. Using the numerical data we obtain $\tau_t \approx 6.2$ ps. The population of level $|2\rangle$ due to thermal back-filling can be completely neglected, due to the high value of $\Delta\mathcal{E}_{inj}$, so that the wall-plug efficiency can be calculated by using Eq. 11.42:

$$\eta_{wp} = \eta_{int} \frac{\alpha_m}{\alpha_m + \alpha_w} \frac{1}{1 + \Delta\mathcal{E}_{inj}/\hbar\omega} \left(1 - \frac{\tau_t}{\tau_{eff}} \frac{\alpha_m + \alpha_w}{n_s N_p \sigma_m}\right) = 0.078,$$

which is very close to the measured value of 7%. In the previous expression the modal gain cross-section σ_m can be calculated by using Eq. 11.6, $\sigma_m = 3.5 \times 10^{-9}$ cm.

11.6 Applications

A major application of QCLs is infrared spectroscopy, where the transmittance or absorbance spectrum of the infrared laser beam is measured after propagation through a sample. Mid-infrared spectral region, in the range between 3 μm and 15 μm is particularly interesting since most chemicals have their fundamental vibrational modes in this spectral region. Chemical-sensing applications range from remote sensing of environmental gases and pollution in the atmosphere, to explosive detection and industrial process monitoring, to medical diagnostics such as breath analysis. For all these applications it is important to use single-mode, narrow-linewidth, tunable mid-infared lasers. This is required to achieve multiple-species detection with high sensitivity and selectivity. This can be accomplished by using QCLs, which naturally emit in the mid-infrared spectral region. Single-mode and tunable operation of QCLs has been demonstrated by using a number of different cavity configurations. For example by using distributed feedback (DFB) QCLs or by employing external cavity QCLs. Just as a recent example, a QCL source has been reported tunable between

6.2 and 9.1 μm obtained by integrating an eigth-laser sampled grating DFB laser array with an on-chip beam combiner[5]. An important feature of the DFB QCL is that its emission wavelength can be continuously tuned by changing the heat sink temperature or the injection current, which changes the temperature of the device. In both cases the effect is a change of the effective refractive index, which determines a shift of the resonance wavelength of the Bragg grating of the DFB structure, with a characteristic tuning rate around 0.1-0.2 $cm^{-1}K^{-1}$. This is particularly important since the DFB QCL can be continuously tuned across the absorption peaks of various molecules, thus offering very high detection sensitivity of trace gases. In order to cover a broad wavelength range multiple DFB QCLs with different grating periods on a single chip can be employed.

Another emerging application of QCLs is free-space communication. The advantage is the possibility to use a wavelength in the highly transmissive windows of atmosphere between 3.3 and 4.2 μm or between 8 and 13 μm. Another important application is the use of QCLs as local oscillators in heterodyne detection systems for the far- and mid-infrared region. Advances in terahertz QCLs have led to a number of applications in this spectral region. Particularly important is the use of QCLs as high-average power laser source for terahertz imaging.

[5]W. Zhou, N. Bandyopadhyay, D. Wu, R. McClintock, M. Razeghi, "Monolithically, widely tunable quantum cascade lasers based on a heterogeneous active region design," Scient. Rep. **6**, 25213 (2016).

A. Fundamental constants

	Symbol	Value
Electron charge	e	$-1.6021766208(98) \times 10^{-19}$ C
Electron rest mass	m_0	$9.10938356(11) \times 10^{-31}$ kg
Planck constant	h	$6.626070040(81) \times 10^{-34}$ J·s
		$4.135667662(25) \times 10^{-15}$ eV·s
Reduced Planck constant	$\hbar$	$1.054571800(13) \times 10^{-34}$ J·s
		$6.582119514(40) \times 10^{-16}$ eV·s
Boltzmann constant	k_B	$1.38064852(79) \times 10^{-23}$ J·K^{-1}
		$8.6173303(50) \times 10^{-5}$ eV·K^{-1}
Light velocity in vacuum	c	2.99792458×10^{8} m·s^{-1}
Vacuum permittivity	ϵ_0	$8.854187817 \times 10^{-12}$ F·m^{-1}
Vacuum permeability	μ_0	$4\pi \times 10^{-7}$ H·m^{-1}
		$1.2566370614 \times 10^{-6}$ H·m^{-1}
Avogadro number	N_A	$6.022140857(74) \times 10^{23}$ mol^{-1}
Bohr radius	a_B	$5.2917721067(12) \times 10^{-11}$ m
Rydberg constant	Ry	$1.097373156 \times 10^{7}$ m^{-1}
		13.605693009(84) eV

Index

A

Absorption coefficient
  bulk semiconductor 94
  intersubband transition 136
  quantum well 133
Acceptor 47
Auger recombination 111

B

Band folding 11
Basis 1
Bernard-Duraffourg condition 98
  quantum well 133
Bimolecular recombination 104
  coefficient 104
Bloch theorem 10
Born-von Karman boundary
    conditions 33
Bragg condition 194
Bravais lattice 1
Brillouin zone 8
Burstein-Moss shift 161

C

Canonical ensemble 72
Canonical momentum 78
Carrier freeze-out 46
Carrier leakage over barrier 165
Carrier overflow 164
Compensated semiconductor 54
Conductivity 53
Confinement factor 125, 179, 180
  enhancement factor 210
  fill factor 210
  VCSEL 210
Coulomb gauge 76
Coupling parameter 198
Crystal momentum 12
Current droop 168

D

de Broglie wavelength 10
Deformation potential
  hydrostatic 143
  shear 143
Degenerate semiconductor 48
Density matrix 69
Density of states 33

Differential gain coefficient 100
Diffusion
coefficient 55
current 55
Einstein relation 55
Diffusion length 157
Direct gap semiconductor 22
Distributed Bragg reflector 212
stop-band 213
Distributed Bragg Reflector laser. 206
Distributed Feedback Laser 193
$\lambda/4$-shifted 195, 205
stop-band 195
uniform grating 201
Donor 43
Doped semiconductor 43
Double heterostructure 160
Drift velocity 52

E

Edge-emitting laser 189
Effective density of states 39
Effective mass 13
conductivity mass 44
quantum well 124
Elastic stiffness tensor 141
Electric dipole approximation 81
Electroluminescence 157
Electron blocking layer 167
Electron capture cross section ... 108
Electron lifetime 110
Electron recycling 219
Electron-phonon scattering 52
Elliott formula
three-dimensions 152
two-dimensions 155
Envelope function 122
approximation 122
overlap integral 130
Exciton 148
absorption spectrum 152
binding energy 149
quantum well 154
radius 149
Expectation value 61
External quantum efficiency 185

F

Fermi energy 38
intrinsic 42
Fermi function 39
Fermi's Golden Rule 67
Fermi-Dirac distribution function.. 37
Fick's law 55
Frenkel exciton 148

G

Gain coefficient 98
quantum well 133
Gain-guided laser 189
Gallium arsenide 29
Gallium nitride 30
Group index 191

H

Heavy-hole band 24
Hermitian operator 60
Heterostructure
graded 167
Hole 14
Hole capture cross-section 109
Hole lifetime 110
Homojunction 157

I

Identity operator 68
Index-guided laser 190
Indirect semiconductor 24
Intersubband transition 131
Intraband transition 131
Intrinsic carrier density 42
Intrinsic semiconductor 41

J

Joint density of states 92
quantum well 132

K

k·p method 25
Kane energy 27

L

Lagrange equations 78
Lagrange undetermined multipliers, method 115
Lagrangian of an electron in an e.m. field 78
Lambertian source 171
Legendre transformation 79
Light Emitting Diode 157
  blue 173
  emission pattern 170
  external efficiency 168
  Internal quantum efficiency . 160
  output power 168
  responsivity 168
  transmission efficiency 168
Light escape cone 169
Light-hole band 24
linear combination of atomic orbitals (LCAO) method 17
Logarithmic losses 179
Longitudinal relaxation time 84
Luminous efficiency 171
Luminous efficiency function 172
Luttinger parameters 126

M

Mass-action law 41, 42
Matthiessen's rule 53
Mean free path 52
Misfit dislocation 140
Mixed state 69
Mobility of an electron 52
Modal gain 180
Modal gain coefficient 180
Multiple quantum well 187

N

Non-degenerate semiconductor ... 40
Nonradiative recombination 107

P

Partition function 69
Photon lifetime 179
Photopic vision 172
Point group 3
Primitive cell 2
Primitive vector 1
Pure state 69

Q

Quantum Cascade Laser 219
  coupling strength 228
  external quantum efficiency . 226
  gain coefficient 221
  internal quantum efficiency . 225
  miniband 220
  minigap 220
  modal cross-section 222
  modal gain 221
  slope efficiency 225
  wall-plug efficiency 226
Quantum efficiency
  internal 119
Quantum well 121
  laser 186
Quasi-Fermi energy 49
Quasi-resonance approximation ... 65

R

Radiative lifetime 100
Radiative recombination time ... 106
Reciprocal lattice 7
Resonant periodic gain 212
Resonant tunneling 226
Rotating phase approximation 65

S

Scattering time 52
Scotopic vision 172
Selection rules 91

quantum well 129
Self-adjoint operator 60
Semi-classical approximation 79
Semiconductor laser
carrier confinement 178
photon confinement 178
rate equations 178
Separate confinement heterostructure 186
Shockley-Read-Hall (SRH) model 107
Silicon 28
Slope efficiency 185
Slowly-varying envelope approximation 197
Sommerfeld enhancement factor 153
Spin-orbit interaction 22
Split-off band 23
Spontaneous
emission 100
transition rate 100
Spontaneous emission factor 179
Spontaneous emission lifetime ... 100
Strain 141
biaxial 142
compressive 141
tensile 141
volumetric 142
Strained quantum wells 139
Stress 140
normal 141
shear 141
tensor 140
Surface recombination 117
velocity 118
Surface state
fast 117
slow 117
Susceptibility 83

T

Thermal back-filling 221
Tight-binding method 15
Time-dependent perturbation theory . 64
Transition probability 65
Transition rate 66
Transparency condition 99
quantum well 133
Transparency density 99
Transverse relaxation time 85
Trap state 107
Truncated inverted pyramid (TIP) 170

U

Unipolar device 219

V

van Roosbroeck-Shockley equation ... 101
Varshni equation 30
Vector potential 75
Vertical Cavity Surface-Emitting Laser 209
von Neumann equation 74

W

wall-plug efficiency
Light Emitting Diode 168
Wannier-Mott exciton 148
Weak-field limit 80
Wigner-Seitz cell 6
Wurtzite 5

Z

Zincblende 5